

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/114310/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Ding, Wenjie ORCID: <https://orcid.org/0000-0003-2774-5777>, Mazouz, Khelifa ORCID: <https://orcid.org/0000-0001-6711-1715> and Wang, Qingwei ORCID: <https://orcid.org/0000-0002-3695-7846> 2019. Investor sentiment and the cross-section of stock returns: New theory and evidence. Review of Quantitative Finance and Accounting 53 , pp. 493-525. 10.1007/s11156-018-0756-z file

Publishers page: <http://dx.doi.org/10.1007/s11156-018-0756-z>  
<<http://dx.doi.org/10.1007/s11156-018-0756-z>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





# Investor sentiment and the cross-section of stock returns: new theory and evidence

Wenjie Ding<sup>1,3</sup> · Khelifa Mazouz<sup>1</sup> · Qingwei Wang<sup>1,2</sup>

© The Author(s) 2018

## Abstract

We extend the noise trader risk model of DeLong et al. (J Polit Econ 98:703–738, 1990) to a model with multiple risky assets to demonstrate the effect of investor sentiment on the cross-section of stock returns. Our model formally demonstrates that market-wide sentiment leads to relatively higher contemporaneous returns and lower subsequent returns for stocks that are more prone to sentiment and difficult to arbitrage. Our extended model is consistent with the existing empirical evidence on the relationship between sentiment and cross-sectional stock returns. Guided by the extended model, we also decompose investor sentiment into long- and short-run components and predict that long-run sentiment negatively associates with the cross-sectional return and short-run sentiment positively varies with the cross-sectional return. Consistent with these predictions, we find a negative relationship between the long-run sentiment component and subsequent stock returns and positive association between the short-run sentiment component and contemporaneous stock returns.

**Keywords** Investor sentiment · Predictive return · Noise trader risk

**JEL Classification** G02 · G11 · G12 · G14

---

We thank two anonymous referees, Cheng-Few Lee (the editor), Darren Duxbury, Danial Hemmings, Dylan C. Thomas as well as participants at Behavioral Finance Working Group Conference, Xiamen-Newcastle-Cardiff Conference and Cardiff University for helpful discussions. All errors and omissions are ours.

---

✉ Khelifa Mazouz  
mazouzk@cardiff.ac.uk

<sup>1</sup> Cardiff Business School, Cardiff, UK

<sup>2</sup> Centre for European Economic Research (ZEW), Mannheim, Germany

<sup>3</sup> Present Address: Shenzhen Audencia Business School, Shenzhen, China

# 1 Introduction

Several theoretical studies, such as Delong et al. (1990) (DSSW hereafter), demonstrate that investor sentiment affects asset prices when rational arbitrageurs face limits to arbitrage.<sup>1</sup> These studies focus on a single risky asset, and accordingly, their models are more suitable for empirical tests involving aggregate market portfolios (Huang et al. 2015). However, while there is ample evidence that market sentiment affects the cross-section of asset returns,<sup>2</sup> a little has been done to explain the theoretical basis for the role of investor sentiment in the context of multiple assets.

In this study, we provide a parsimonious and tractable model of how investor sentiment affects the cross-section of stock returns. We extend the DSSW model by introducing multiple risky assets that differ in their exposure to market-wide sentiment. Our analysis is motivated by the premise that predictions from a single-asset model do not necessarily hold in multi-asset economies (Verrecchia 2001). For example, Cochrane et al. (2008) show that price-dividend ratio is constant in the one-tree model of Lucas (1978), but varies over time in a two-tree model. Therefore, it is unclear whether DSSW predictions can be generalized to markets with more than one risky asset.

To shed some light on this issue, we develop a simple model that formalizes Baker and Wurgler's (2006) idea that sentiment-prone assets are also more difficult to arbitrage and provides the theoretical intuition for the widely documented evidence that investor sentiment affects the cross-sectional asset returns. Our model assumes that there are two types of risky assets A and B and that irrational investors' beliefs are biased upwards (downwards) more towards A than B when market sentiment is high (low). That is to say, asset A has higher exposure to market-wide sentiment (more sentiment-prone) than asset B. When investor sentiment is uncertain, this assumption also implies that the equilibrium returns of asset A relative to the returns of asset B will fluctuate more with the shift in market sentiment, and hence posing higher noise trader risk to rational arbitrageurs. Thus, the contemporaneous returns of asset A are expected to exhibit greater sensitivity to the changes in investor sentiment than the contemporaneous returns of asset B. The returns of asset A are also expected to reverse more than the returns of asset B as investor sentiment eventually reverts to its mean. Therefore, the return difference between the more sentiment-prone asset and the less sentiment-prone asset should be positively associated with the change in contemporaneous sentiment and negatively related to the level of lagged sentiment. These predictions are in line with the existing empirical evidence on the relationship between sentiment and cross-sectional stock returns.

Similar to DSSW, our model of sentiment effect features long- and the short-run investor sentiment components. The long-run sentiment reflects the average bullishness of noise traders, while the short-run sentiment represents the transitory deviations from the long-run sentiment. Both components affect the price of the single risky asset in the DSSW model. Unlike DSSW, the two components in our model have cross-sectional implications. When the long-run sentiment remains the same and the short-run component increases,

<sup>1</sup> Several other models, including Campbell and Kyle (1993), Daniel et al. (1998), Barberis et al. (1998), Hirshleifer (2001), also illustrate the effect of sentiment on signal asset returns.

<sup>2</sup> For example, Brown and Cliff (2004), Brown and Cliff (2005), Baker and Wurgler (2006), Lemmon and Portniaguina (2006), Qiu and Welch (2004), Kumar and Lee (2006), Frazzini and Lamont (2008), Li and Yeh (2011), Stambaugh et al. (2012), Ben-Rephael et al. (2012), Da et al. (2015), Huang et al. (2015), among many others.

irrational investors become more bullish and drive up the relative returns of more sentiment-prone stocks. Hence, we predict a positive correlation between contemporaneous changes in the short-run sentiment and the relative returns of sentiment-prone stocks over sentiment-immune stocks. However, a higher long-run sentiment exerts more upward pressure on the prices of more sentiment prone stocks and reduce their expected future returns.<sup>3</sup> Therefore, we predict the long-run component to be a contrarian predictor of subsequent cross-sectional returns.

Motivated by our theoretical model, we empirically decompose investor sentiment into a short-run sentiment component, constructed as incremental changes of sentiment, and a long-run sentiment component, measured by a moving average of investor sentiment in the past 2 years.<sup>4</sup> We test the cross-sectional pricing effect of both components with common stock listed on NYSE, AMEX, and NASDAQ between July 1965 and Sep 2015. We follow Baker and Wurgler (2006) and construct sixteen long-short portfolios that buy more sentiment-prone stocks and short less sentiment-prone stocks. Sentiment-prone stocks tend to be small, young, more volatile, unprofitable, non-dividend-paying, and distressed, with extreme growth potential or with a relatively high percentage of intangible assets.

Consistent with our theoretical predictions, we find a positive relationship between short-run sentiment and contemporaneous cross-sectional stock returns and a negative association between long-run sentiment component and subsequent cross-sectional stock returns. Further analysis suggests that the effect of the sentiment components on returns is generally stronger for stocks that are small, young, volatile, unprofitable/non-dividend paying, financially distressed and have high growth potential. These findings are robust after accounting for systematic risk and time-varying factor loadings as well as to alternative sentiment measures, to alternative constructions of portfolios, and to alternative decomposition of sentiment. Additionally, the effect of decomposed investor sentiment remains strong after taking the effect of investor attention into account.

The contribution of our paper is twofold. First, we contribute to the theory of investor sentiment by presenting a parsimonious model that explicitly examines the effect of market-wide sentiment on the cross-sectional asset returns. Such a model provides a theoretical underpinning for the existing empirical evidence that market-wide sentiment does not have a uniform impact on stock returns. It also captures well the idea of Baker and Wurgler (2006) that more sentiment-prone assets are also more difficult to arbitrage. The type of limits to arbitrage we consider here is the noise trader risk, while limits to arbitrage in Baker and Wurgler (2006) take many other forms, including transaction costs and idiosyncratic risk. Admittedly, our multiple-assets version of noise trader risk model is a simple and straightforward extension of DSSW (1990), and is not intended to model the belief formation process that originates from the behavioral heuristic evidenced in psychology literature (Barberis et al. 1998) or sentiment that arises endogenously. Instead, our simple model aims to bridge the gap between the theoretical studies, which focus on the role of sentiment in context of a single risky asset, and empirical literature, which provides ample evidence on the cross-sectional pricing impact of investor sentiment.

<sup>3</sup> We do not model time-varying long-run sentiment explicitly. Instead, we rely on comparative statics to obtain predictions on the pricing effect of long-run sentiment. Allowing the time varying long-run sentiment complicates the model although we expect the effect of long-run sentiment to remains the same.

<sup>4</sup> We also consider short-run and long-run sentiment components constructed from 12-month, 36-month, and 48-month window period and find similar results.

Second, we contribute to the empirical literature on investor sentiment by decomposing investor sentiment into short- and long-run components and show that both components affect cross-sectional stock returns. Existing empirical studies on the pricing impact of investor sentiment find that change in investor sentiment is positively associated with contemporaneous returns (e.g., Lee et al. 2002; Brown and Cliff 2004; Ben-Rephael et al. 2012) and the sentiment level is negatively related to future returns in the short run (e.g. Brown and Cliff 2004; Baker and Wurgler 2006; Lemmon and Portniaguina 2006). Only a few papers investigate the long-run sentiment-return relationship. For example, Brown and Cliff (2005) document that investor sentiment level is a significant predictor of long-run returns up to 36-month horizon in size and value portfolios, and Schmeling (2009) shows that consumer confidence index predicts aggregate market returns and the cross-sectional stock returns up to a 24-month forecast horizon in 18 industrialized countries. Chen and Kuo (2014) find the stronger transitory effect of investor sentiment on interest rate smile. However, these studies do not decompose sentiment into long- and short-run components and, therefore, do not attempt to separate the effect of the two sentiment components on the cross-sectional stock returns. Unlike prior studies, our empirical analysis is guided by a theoretical model, which provides explicit predictions on how long- and short-run components of investor sentiment would affect cross-sectional stock returns. Consistent with the theoretical predictions, we find that stock returns are negatively associated with the long-run sentiment component and positively related to the short-run sentiment component.

The rest of this paper is organized as follows. Section II illustrates the model and derives the two main hypotheses. Section III describes the data. Section IV discusses the empirical results and the robustness checks. Section V stresses the conclusions.

## 2 A cross-sectional noise trader risk model

DSSW (1990) propose a simple overlapping generation model of a market with one risky asset and one risk-free asset and two types of two-period-lived agents, sophisticated investors and irrational noise traders with stochastic misperception. The uncertainty of noise traders' misperception creates "noise trader risk" that deters rational investors from fully arbitraging. Because of its single risky asset setting, DSSW model is presumably better suited for explaining the impact of investor sentiment at the aggregate levels (Huang et al. 2015). Since the claims or results in a single-asset model can sometimes be reversed in multi-asset economies (Verrecchia 2001), it is unclear whether predictions of DSSW can be generalized to markets with more than one risky asset. In this study, we extend the single risky asset model of DSSW to a noise trader risk model with multiple risky assets that vary in their exposure to market-wide investor sentiment.

In the DSSW noise trader risk model, sophisticated investors (denoted as "*i*") have rational In the DSSW noise trader risk model, sophisticated investors (denoted as "*i*") have rational expectations, and noise traders (denoted as "*n*") hold biased belief and trade on noise. The percentage of noise traders in the market is set as  $\mu$ , and the percentage of sophisticated investors is  $1 - \mu$ . Both noise traders and sophisticated investors' utility function is a constant absolute risk aversion (CARA) function of wealth,  $U = -e^{-(2\gamma)\omega}$ , where  $\gamma$  is the coefficient of absolute risk aversion and  $\omega$  is wealth. If holding period returns are normally distributed, solving expected utility optimization is equivalent to maximizing  $\bar{\omega} - \gamma\sigma_{\omega}^2$ , where  $\bar{\omega}$  is the expected final wealth, and  $\sigma_{\omega}^2$  is the one period ahead variance of the expected wealth.



There are two assets in the DSSW model, one risky asset and one risk-free asset. The difference between the risk-free asset and the unsafe assets lies in their supply. The risk-free asset is in perfectly elastic supply, which implies that its price is fixed. However, the supply of the unsafe asset is fixed at one unit, which means that its prices fluctuate along with the change in demand. In each period the risk-free asset has a fixed real rate of  $r$  and the risky assets have fixed dividend rate  $r$ .

Unlike DSSW, our model has two risky assets, assets  $u_1$  and  $u_2$ , in addition to the risk-free asset.<sup>5</sup> Sophisticated investors choose to hold  $\lambda_{t,1}^i$  of the risky asset  $u_1$  and  $\lambda_{t,2}^i$  of the risky asset  $u_2$  to maximize their expected utility. However, given their misperception, noise traders maximize their expected utility by choosing a portfolio with holding  $\lambda_{t,1}^n$  of risky asset  $u_1$  and holding  $\lambda_{t,2}^n$  of risky asset  $u_2$ . Due to the various types of limits of arbitrage described in Shleifer and Vishny (1997), sophisticated and noisy traders cannot make unlimited bidding against each other.

Assume that the overall market sentiment  $\rho_t$  follows a normal distribution with  $\rho_t \sim N(\rho^*, \sigma_\rho^2)$ . To examine the cross-sectional effect of investor sentiment and noise trader risk, we further assume that noise traders have different misperceptions of the risky assets  $u_1$  and  $u_2$ .

$$\rho_{t,1} = \alpha_1 \rho_t + \varepsilon_{t,1} \quad (1)$$

$$\rho_{t,2} = \alpha_2 \rho_t + \varepsilon_{t,2} \quad (2)$$

$$\varepsilon_{t,1} \sim N(0, \sigma_{\varepsilon_1}^2), \varepsilon_{t,2} \sim N(0, \sigma_{\varepsilon_2}^2) \quad (3)$$

$$\text{cov}(\varepsilon_{t,1}, \rho_t) = 0, \text{cov}(\varepsilon_{t,2}, \rho_t) = 0, \text{cov}(\varepsilon_{t,1}, \varepsilon_{t,2}) = 0 \quad (4)$$

Equations (1) and (2) show that noise traders' misperception of a risky asset contains a systematic component proportional to market sentiment and an idiosyncratic component. The new disturbance terms that we added to the DSSW's model are the idiosyncratic investor sentiment,  $\varepsilon_{t,i}$ . Such idiosyncratic disturbances capture the firm-specific investor sentiment that is often observed in the financial markets. Our focus, however, is not the pricing role of the idiosyncratic investor sentiment, but the pricing effect of market sentiment,  $\rho_t$ .

While there is no theory supporting a particular stochastic structure of investor sentiment, Eqs. (1) and (2) can be motivated in three ways: first, we do observe episodes of bull and bear market and associated waves of market-wide sentiment (e.g., during the bubble period in American stocks in the 1920s), as well as sentiment towards individual stocks (e.g., the South Sea company bubble). Second, while existing empirical studies mainly focus on market-wide sentiment, there is a growing literature on the role of idiosyncratic sentiment, often measured by the degrees of positive and negative tones in a firm's media coverage, corporate disclosures, internet posts, and overnight returns (e.g., Tetlock 2007; Bhattacharya et al. 2009; Kim and Kim 2014; Aboody et al. 2018). Finally, some theoretical models assume that future dividends or firm profitability are linear in the market and firm-specific factors (e.g., Peng and Xiong 2006; Jackson et al. 2017).<sup>6</sup> If noise traders

<sup>5</sup> Extending the model further with more than two risky assets is straightforward. We focus on the two risky assets model since it is sufficient to shed the lights on the cross-sectional effect of investor sentiment.

<sup>6</sup> A number of empirical studies document that market and firm-specific factors help explain analyst forecast accuracy (e.g., Hutton et al. 2012). Practitioners also place great emphasis on the market factor when estimating a firm's future profitability (Jackson et al. 2018).

form their (biased) beliefs about future price on the basis of these models, then their beliefs are likely to be linearly related to market and firm-specific misperceptions.

Equations (1)–(4) above imply that  $\sigma_{\rho_1}^2 = \alpha_1^2 \sigma_\rho^2 + \sigma_{\varepsilon_1}^2$  and  $\sigma_{\rho_2}^2 = \alpha_2^2 \sigma_\rho^2 + \sigma_{\varepsilon_2}^2$ . For simplicity, we also assume  $\sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2$ .<sup>7</sup> Without the loss of generality, assume  $\alpha_1 > \alpha_2 > 0$ , then  $u_1$  has larger exposure to market sentiment than asset  $u_2$ .<sup>8</sup> This implies  $\sigma_{\rho_1}^2 > \sigma_{\rho_2}^2$ . Latter we show that equilibrium price volatility of asset  $u_1$  is larger than that of asset  $u_2$  due to higher noise trader risk ( $\sigma_{\rho_1}^2 > \sigma_{\rho_2}^2$ ). Higher noise trader risk poses stronger limits to arbitrage for rational investors to trade against irrational investors. As a result, our model parsimoniously captures the intuitive observation of Baker and Wurgler (2006) that more sentiment-prone assets are also more difficult to arbitrage, although the limits to arbitrage in Baker and Wurgler (2006) are broader and not necessarily related to noise trader risk.

For sophisticated investors maximization of their expected utility is equivalent to maximize

$$\begin{aligned} \overline{w^i} - \gamma \sigma_{w^i}^2 = & c_0 + \lambda_{i,1}^i (r + {}_t p_{t+1,1} - p_{t,1}(1+r)) + \lambda_{i,2}^i (r + {}_t p_{t+1,2} - p_{t,2}(1+r)) \\ & - \gamma \left[ \lambda_{i,1}^{i^2} {}_t \sigma_{p_{t+1,1}}^2 + \lambda_{i,2}^{i^2} {}_t \sigma_{p_{t+1,2}}^2 + 2\lambda_{i,1}^i \lambda_{i,2}^i {}_t \text{cov}(p_{t+1,1}, p_{t+1,2}) \right] \end{aligned} \quad (5)$$

For noise traders maximization of their expected utility is equivalent to maximize

$$\begin{aligned} \overline{w^n} - \gamma \sigma_{w^n}^2 = & c_0 + \lambda_{i,1}^n (r + {}_t p_{t+1,1} - p_{t,1}(1+r)) + \lambda_{i,2}^n (r + {}_t p_{t+1,2} - p_{t,2}(1+r)) \\ & - \gamma \left[ \lambda_{i,1}^{n^2} {}_t \sigma_{p_{t+1,1}}^2 + \lambda_{i,2}^{n^2} {}_t \sigma_{p_{t+1,2}}^2 + 2\lambda_{i,1}^n \lambda_{i,2}^n {}_t \text{cov}(p_{t+1,1}, p_{t+1,2}) \right] \\ & + \lambda_{i,1}^n (\alpha_1 \rho_t + \varepsilon_{t,1}) + \lambda_{i,2}^n (\alpha_2 \rho_t + \varepsilon_{t,2}) \end{aligned} \quad (6)$$

where  ${}_t \sigma_{p_{t+1,1}}^2$  is the conditional expectation of one-step-ahead variance of  $p_{t+1,1}$ ,  ${}_t \sigma_{p_{t+1,2}}^2$  is the conditional expectation of one-step-ahead variance of  $p_{t+1,2}$ , and  ${}_t \text{cov}(p_{t+1,1}, p_{t+1,2})$  is the conditional expectation of the covariance of the one-step-ahead risky assets' price  $p_{t+1,1}$  and  $p_{t+1,2}$ . Solving the above optimization problem with first order conditions yields the portfolio holdings of the two risky assets:

$$\lambda_{i,1}^i = \frac{kR_{t+1,2} - \sigma_2^2 R_{t+1,1}}{2\gamma(k^2 - \sigma_1^2 \sigma_2^2)} \quad (7)$$

$$\lambda_{i,2}^i = \frac{kR_{t+1,1} - \sigma_1^2 R_{t+1,2}}{2\gamma(k^2 - \sigma_1^2 \sigma_2^2)} \quad (8)$$

$$\lambda_{i,1}^n = \frac{k(R_{t+1,2} + \alpha_2 \rho_t + \varepsilon_{t,2}) - \sigma_2^2 (R_{t+1,1} + \alpha_1 \rho_t + \varepsilon_{t,1})}{2\gamma(k^2 - \sigma_1^2 \sigma_2^2)} \quad (9)$$

<sup>7</sup> The proposition derived from this extended model will be the same if  $\sigma_{\varepsilon_1}^2 > \sigma_{\varepsilon_2}^2$ . Generally speaking, we consider the idiosyncratic sentiment to have a much smaller sample variance compared with the systematic sentiment component. Our conclusion do not rely on the relationship between  $\sigma_{\varepsilon_1}^2$  and  $\sigma_{\varepsilon_2}^2$ .

<sup>8</sup> We relax the assumption of positive  $\alpha_1$  and  $\alpha_2$  in the discussions at the end of this section.

$$\lambda_{t,2}^n = \frac{k(R_{t+1,1} + \alpha_1 \rho_t + \varepsilon_{t,1}) - \sigma_1^2(R_{t+1,2} + \alpha_2 \rho_t + \varepsilon_{t,2})}{2\gamma(k^2 - \sigma_1^2 \sigma_2^2)} \quad (10)$$

where  $R_{t+1,1} = r + {}_t p_{t+1,1} - p_{t,1}(1+r)$ ,  $R_{t+1,2} = r + {}_t p_{t+1,2} - p_{t,2}(1+r)$ ,  $k = {}_t \text{cov}(p_{t+1,1}, p_{t+1,2})$ , and  $\sigma_1^2 = {}_t \sigma_{p_{t+1,1}}^2$ ,  $\sigma_2^2 = {}_t \sigma_{p_{t+1,2}}^2$ .

Market clearing requires the holding of the two risky assets from the noise traders and the sophisticated investors to be equal to their supply.

$$\begin{cases} (1-\mu)\lambda_{t,1}^i + \mu\lambda_{t,1}^n = 1 \\ (1-\mu)\lambda_{t,2}^i + \mu\lambda_{t,2}^n = 1 \end{cases}$$

By assuming that equilibrium prices in all periods have identical distributions, we can obtain the equilibrium pricing functions by solving the following function recursively:

$$p_{t,1} = \frac{1}{1+r} [r + {}_t p_{t+1,1} - 2\gamma(k + \sigma_1^2) + \mu(\alpha_1 \rho_t + \varepsilon_{t,1})]$$

$$p_{t,2} = \frac{1}{1+r} [r + {}_t p_{t+1,2} - 2\gamma(k + \sigma_2^2) + \mu(\alpha_2 \rho_t + \varepsilon_{t,2})]$$

If the conditional variance of the price is constant, substituting the conditional one-step-ahead price forward yields:

$$p_{t,1} = 1 + \frac{\mu\alpha_1(\rho_t - \rho^*)}{1+r} + \frac{\mu\alpha_1\rho^*}{r} - \frac{2\gamma(k + \sigma_1^2)}{r} + \frac{\mu\varepsilon_{t,1}}{1+r} \quad (11)$$

$$p_{t,2} = 1 + \frac{\mu\alpha_2(\rho_t - \rho^*)}{1+r} + \frac{\mu\alpha_2\rho^*}{r} - \frac{2\gamma(k + \sigma_2^2)}{r} + \frac{\mu\varepsilon_{t,2}}{1+r} \quad (12)$$

The equilibrium price is a function of both the misperception level and the noise trader risk induced by stochastic misperception. The second term in the equilibrium price function shows the change in the price caused by the fluctuations of the overall market misperceptions around its long-run mean. When noise traders become more bullish relative to the average overall market misperceptions, their demand pushes the price up. The third term captures the deviation of the price from the fundamental value caused by the average overall market misperceptions. The fourth term captures the compensation for bearing the “noise trader risk” created by the uncertainty of the next period’s misperception. Noise trader risk makes sophisticated investors unwilling to trade entirely against noise traders since future misperceptions of noise traders can become extreme. The type of limits to arbitrage that deter sophisticated investor from trading against noise relates mainly to the uncertainty about the idiosyncratic and systematic components of sentiments.<sup>9</sup> The last term captures the fluctuation in price caused by the variation of idiosyncratic misperception towards risky asset  $u_1$ . Equations (11) and (12) imply that the unconditional price volatility and the price covariance of the two risky assets at time  $t+1$  are given as:

$$\sigma_{t+1,1}^2 = \frac{\alpha_1^2 \mu^2 \sigma_\rho^2}{(1+r)^2} + \frac{\mu^2 \sigma_{\varepsilon_1}^2}{(1+r)^2} \quad (13)$$

<sup>9</sup> Limits to arbitrage in Baker and Wurgler (2006) take many other forms, such as transaction costs and idiosyncratic risk.



$$\sigma_{t+1,2}^2 = \frac{\alpha_2^2 \mu^2 \sigma_\rho^2}{(1+r)^2} + \frac{\mu^2 \sigma_{\varepsilon_2}^2}{(1+r)^2} \quad (14)$$

$$\text{cov}(p_{t+1,1}, p_{t+1,2}) = \frac{\alpha_1 \alpha_2 \mu^2 \sigma_\rho^2}{(1+r)^2} \quad (15)$$

We also solve the Eqs. (7)–(10) to obtain the portfolio holdings for sophisticated investors and noise traders:

$$\lambda_{t,1}^i = 1 - \frac{\mu(k\alpha_2 - \sigma_2^2\alpha_1)}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)}\rho_t - \frac{\mu(k\varepsilon_{t,2} - \sigma_2^2\varepsilon_{t,1})}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)} \quad (16)$$

$$\lambda_{t,2}^i = 1 - \frac{\mu(k\alpha_1 - \sigma_1^2\alpha_2)}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)}\rho_t - \frac{\mu(k\varepsilon_{t,1} - \sigma_1^2\varepsilon_{t,2})}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)} \quad (17)$$

$$\lambda_{t,1}^n = 1 + \frac{(1-\mu)(k\alpha_2 - \sigma_2^2\alpha_1)}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)}\rho_t + \frac{(1-\mu)(k\varepsilon_{t,2} - \sigma_2^2\varepsilon_{t,1})}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)} \quad (18)$$

$$\lambda_{t,2}^n = 1 + \frac{(1-\mu)(k\alpha_1 - \sigma_1^2\alpha_2)}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)}\rho_t + \frac{(1-\mu)(k\varepsilon_{t,1} - \sigma_1^2\varepsilon_{t,2})}{2\gamma(k^2 - \sigma_1^2\sigma_2^2)} \quad (19)$$

Given the obtained price volatility and covariance, then  $2\gamma(k^2 - \sigma_1^2\sigma_2^2) < 0$ ,  $k\alpha_2 - \sigma_2^2\alpha_1 < 0$ , and  $k\alpha_1 - \sigma_1^2\alpha_2 < 0$ . Thus, the sophisticated investors' holdings of the two risky assets are inversely proportional to current market sentiment, while the noise traders' holdings of these risky assets are positively associated with the current market sentiment. That is, sophisticated investors reduce their demand for sentiment-prone assets while noise traders increase their demand for sentiment-prone assets when overall market sentiment becomes more bullish.

The exposure of risky assets to the overall market misperception also affects the portfolio holdings. The sophisticated investors' holdings of  $u_1(u_2)$  is positively related to  $\alpha_1(\alpha_2)$ . Since  $\alpha_1 > \alpha_2$ , the current market sentiment will have a greater effect on investors' (both sophisticated and irrational) holdings of asset  $u_1$ . This also means that when overall market sentiment becomes more bullish, sophisticated investors will reduce their demand for asset  $u_1$  more than asset  $u_2$  and irrational investors will increase their demand for asset  $u_1$  more than on asset  $u_2$ .

Recall that the excess return from date  $t$  to date  $t+1$  is noted as:  $R_{t+1} = r + {}_t p_{t+1} - p_t(1+r)$ . Plugging it in the equilibrium price, we obtain the return for asset  $u_1$  and asset  $u_2$  at time  $t+1$

$$R_{t+1,1} = \frac{\mu\alpha_1[\rho_{t+1} - (1+r)\rho_t]}{1+r} - \frac{\alpha_1\mu\rho^*}{1+r} + 2\gamma(k + \sigma_1^2) + \theta_1 \quad (20)$$

$$R_{t+1,2} = \frac{\mu\alpha_2[\rho_{t+1} - (1+r)\rho_t]}{1+r} - \frac{\alpha_2\mu\rho^*}{1+r} + 2\gamma(k + \sigma_2^2) + \theta_2 \quad (21)$$

where  $\theta_1$  ( $\theta_2$ ) are functions of  $\varepsilon_{t,1}$  and  $\varepsilon_{t+1,1}$  ( $\varepsilon_{t,2}$  and  $\varepsilon_{t+1,2}$ ). Hence, the return difference between the two risky assets is

$$R_{t+1,1} - R_{t+1,2} = (\alpha_1 - \alpha_2) \left[ \frac{\mu[\rho_{t+1} - (1+r)\rho_t]}{1+r} - \frac{\mu\rho^*}{1+r} \right] + 2\gamma(\sigma_1^2 - \sigma_2^2) + \theta_1 - \theta_2 \quad (22)$$

The equilibrium returns can also be expressed as functions of the deviation of current sentiment from its long-run mean, denoted as  $\eta_t$  ( $\eta_t = \rho_t - \rho^*$ ).

$$R_{t+1,1} = \frac{\mu\alpha_1[\eta_{t+1} - (1+r)\eta_t]}{1+r} - \alpha_1\mu\rho^* + 2\gamma(k + \sigma_1^2) + \theta_1 \quad (23)$$

$$R_{t+1,2} = \frac{\mu\alpha_2[\eta_{t+1} - (1+r)\eta_t]}{1+r} - \alpha_2\mu\rho^* + 2\gamma(k + \sigma_2^2) + \theta_2 \quad (24)$$

Hence, the return difference between asset  $u_1$  and asset  $u_2$  at time  $t+1$  can be written as

$$R_{t+1,1} - R_{t+1,2} = (\alpha_1 - \alpha_2) \left[ \frac{\mu[\eta_{t+1} - (1+r)\eta_t]}{1+r} - \mu\rho^* \right] + 2\gamma(\sigma_1^2 - \sigma_2^2) + \theta_1 - \theta_2 \quad (25)$$

Take the unconditional expectations of Eq. (25), we obtain:

$$E(R_{t+1,1} - R_{t+1,2}) = (\alpha_1 - \alpha_2) [-\mu\rho^*] + 2\gamma(\sigma_1^2 - \sigma_2^2) \quad (26)$$

We consider  $\rho^*$ , the mean of the overall market misperception, as the long-run sentiment component, and the incremental sentiment change as the short-run sentiment component. The latter is measured by either sentiment increment,  $\rho_t - (1+r)\rho_{t-1}$ , or the changes of sentiment's derivation from its long run mean,  $\eta_{t+1} - (1+r)\eta_t$ .

Equations (20) and (21) show that the short-run sentiment is positively related to the returns of both risky assets. The effect is stronger for asset  $u_1$  than  $u_2$  because  $\alpha_1 > \alpha_2$  [see Eq. (25)]. As a result, we have the following hypothesis on the pricing effect of the short-run component of investor sentiment:

**Hypothesis 1** The short-run sentiment component is positively related to the contemporaneous difference in returns of more sentiment-prone and less sentiment-prone assets.

Inspection of Eqs. (20) and (21) also reveals that returns of a risky asset depend on the exposure of the misperception about its price to the overall market misperception. High long-run sentiment implies lower equilibrium returns for both risky assets ( $\alpha_1$  and  $\alpha_2$  are positive), and the returns of  $u_1$  are more adversely affected by the long run sentiment than  $u_2$  ( $\alpha_1 > \alpha_2$ ). This leads to our hypothesis on the pricing effect of the long-run component of investor sentiment.

**Hypothesis 2** The long-run sentiment component predicts lower returns of more sentiment-prone assets relative to returns of less sentiment-prone assets.

If we relax the assumption that  $\alpha_1 > \alpha_2 > 0$  by allowing  $\alpha_2 < 0$  while keeping  $\alpha_1 > 0$ , the effect of long- and short-run investor sentiment are inverted for the risky asset  $u_2$ . In this case, when the average market sentiment becomes more bullish, it exerts a negative effect

on contemporaneous returns due to the short-run sentiment effect, and a positive effect on future returns due to the long-run sentiment effect. The opposite is true in case of the risky asset  $u_1$ , as it still has a positive exposure to market sentiment. Hence, the returns of the two risky assets move in the opposite directions, and the effect of long- and short-run sentiment will be muted at the aggregate market level. Baker and Wurgler (2007) make a similar argument that if the price of a low sentiment-prone stock is inversely related to sentiment, due to reasons such as “flight to quality”, then the effect of sentiment on the aggregate market return is unlikely to be significant. This implication may also explain the inconclusive evidence on the impact of investor sentiment on aggregate market returns (Chen et al. 1993; Elton et al. 1998; Brown and Cliff 2004; Baker and Wurgler 2007; Baker et al. 2012). However, for a given positive  $\alpha_1$  when  $\alpha_1 > \alpha_2$ , a negative (instead of positive)  $\alpha_2$  makes the cross-sectional returns difference between  $u_1$  and  $u_2$  more dispersed. Hence, it is not surprising that several empirical studies find that market sentiment affects asset returns in the cross-section (Baker and Wurgler 2006; Lemmon and Portniaguina 2006; among others).

### 3 Data

#### 3.1 Portfolio construction

To test our theory, we rely on Baker and Wurgler (2006) to find out the proxies related to how *sensitive* a stock is to the overall market sentiment (i.e.,  $\alpha_1$  and  $\alpha_2$  in our model) and use the Baker–Wurgler sentiment index as our baseline measure of market sentiment (i.e.,  $\rho_t$  in our model). Baker and Wurgler (2006) argue that firms that are small, young, volatile, non-dividend-paying, non-profitable, informationally opaque, financially distressed or have more growth opportunity are more sentiment-prone. Similar to Baker and Wurgler (2006), we construct sixteen long-short portfolios based on ten characteristics representing the sentiment-prone level of a stock. The ten characteristics are market capitalization (ME), firm age (Age), total risk (Sigma), earnings-book ratio for profitable firms (E/BE), dividend-book ratio for dividend payers (D/BE), fixed assets ratio (PPE/A), research and development ratio (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A) and sales growth ratio (GS).<sup>10</sup> These proxies could measure the exposure to systematic sentiment in our model because they reflect how attractive a stock is to speculative demand (and how difficult it is for sophisticated investor to arbitrage mispricing).<sup>11</sup>

The monthly stock returns are from CRSP. The stock market data include the returns of all common stocks (share codes in 10 and 11) in NYSE, AMEX, and NASDAQ (with stock exchange code in 1 2 3) between January 1962 and Sep 2015. Our final sample consists of 18,175 firms. The firm-level accounting data are from Compustat. We adopt the Fama and French (1992) approach and match the year-end accounting data of year  $t-1$  to monthly returns from July  $t$  to June  $t+1$ . The ten firm characteristics are winsorized at 99.5 and 0.5% annually. The breakpoints for deciles are defined only using NYSE firms. The

<sup>10</sup> Details on these characteristics variables are provided in Table 7 in the “Appendix”.

<sup>11</sup> Several studies also support that these ten characteristics measures the elasticity of a stock towards investor sentiment. Berger and Turtle (2012) test whether the ten characteristics measures sentiment-prone level effectively by calculating the average sentiment beta in the cross-sectional decile portfolios sorted on these characteristics. They conclude that investor sentiment sensitivities are significantly correlated with their categorization measures. Sharma et al. (2015) find that time-varying herding behaviour is more prevalent in certain stock sectors.

top three, middle four and bottom three decile portfolios used to construct the long-short portfolio, are denoted as H, M and L, respectively. We follow Baker and Wurgler's (2006) approach to construct the long-short portfolios and obtain the relative returns of more sentiment-prone stocks over less sentiment-prone stocks when possible.<sup>12</sup>

To facilitate the interpretation of regression results, we use the returns of more sentiment-prone stocks minus the returns of less sentiment-prone stocks as the dependent variables. For example,  $BE/ME(L-M)$  represents the return difference between the bottom three deciles and the middle four deciles when using  $BE/ME$  as the sorting characteristic. All the portfolio returns are equal-weighted.<sup>13</sup>

The sixteen long-short portfolios can be categorized into six groups. The first group is named "Size, Age, and Risk". Small, young, and volatile firms are sentiment-prone, and therefore the long-short portfolios constructed on  $ME$ , Age and Sigma are denoted as  $ME(L-H)$ ,  $Age(L-H)$ , and  $Sigma(H-L)$ , respectively. The second group is referred to as "Profitability and Dividend Policy". In this group, non-profitable stocks or stocks with low dividend payments are considered to be more prone to sentiment. The third group is labelled "Tangibility" and contains portfolios constructed by  $PPE/A$  and  $RD/A$ . Stocks with more intangible assets are considered to be more informationally opaque and therefore more prone to sentiment. The last three groups are called "Growth Opportunity and Distress", "Growth Opportunities" and "Distress", respectively, and consist of portfolios that are constructed from  $BE/ME$ ,  $EF/A$ , and  $GS$ .

The reason for constructing nine long-short portfolios based on the last three characteristics ( $BE/ME$ ,  $EF/A$ , and  $GS$ ) is that those three characteristics have a multidimensional nature. Stocks at the top and the bottom deciles sorted on  $BE/ME$ ,  $EF/A$  or  $GS$  represent either extreme growth or extreme value stocks, while firms in the middle deciles are more stable and safer. In the meantime, those three characteristics are also seen as a generic pricing factor in literature. Take  $BE/ME$  as an example. High  $BE/ME$  implies that a firm is in distress, while low  $BE/ME$  indicates extreme growth potential. On one hand, financially distressed stocks are highly appealing to the speculative demand of irrational investors, firms with high  $BE/ME$ , as low  $EF/A$  or low  $GS$  are likely to be more prone to sentiment. On the other hand, as high growth firms are hard to value, the returns of firms with low  $BE/ME$ , high  $EF/A$ , and high  $GS$  are likely to be affected more by shifts in investor sentiment.

Panel A of Table 1 presents the summary statistics for the long-short portfolio returns. Most of the long-short portfolios have positive average returns. The negative average portfolio returns of  $EF/A(H-L)$  and  $GS(H-L)$  are not surprising, because  $EF/A$  and  $GS$  measure the growth opportunity and financial stress in the opposite way from  $BE/ME$ .<sup>14</sup> Returns of the sixteen long-short portfolios are all positively skewed. The last two columns in Panel A report the first-order autoregression coefficients ( $AR(1)$ ) and the correlation between the variable and 1-month lagged Baker–Wurgler sentiment ( $Corr$ ), respectively. Although the

<sup>12</sup> We only consider the  $RD/A$  from 1972 because the  $RD/A$  data are not available until 1971. We use the same variable definition of Baker and Wurgler (2006) except for  $RD/A$ , which we do not replace any missing value with zero. Replacing R&D missing values with zeros will cause some problem after mid 2000 where almost half of the observations are zero. We focus on the stocks that have non-missing R&D values. Monthly returns are adjusted for delisting.

<sup>13</sup> We also conduct tests with value-weighted returns and find similar results.

<sup>14</sup> One may find it more appropriate to construct long-short portfolios as  $EF/A(L-H)$  and  $GS(L-H)$ . However, due to the multidimensional nature of  $EF/A$  and  $GS$ , we would not expect decomposed investor sentiment to perform well in predicting long-short portfolio returns calculated with top and bottom portfolio of  $EF/A$  or  $GS$ , whatever the long-legs or short-legs are. Accordingly, we use  $EF/A(H-L)$  and  $GS(H-L)$  to be consistent with Baker and Wurgler (2006).

**Table 1** Summary statistics

Variable		Obs	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	AR(1)	Corr
<i>Panel A Dependent variables</i>										
ME	L–H	606	0.28	3.43	–12.73	17.10	0.78	6.29	0.08	–0.14***
Age	L–H	606	0.07	3.34	–12.69	23.89	1.23	10.03	0.15*	–0.17***
Sigma	H–L	606	0.12	4.48	–13.88	24.26	0.76	6.03	0.22***	–0.18***
E	<0–>0	606	0.13	4.06	–14.44	28.50	1.36	9.60	0.18**	–0.18***
D	=0–>0	606	0.10	3.73	–13.46	24.39	1.03	8.74	0.19**	–0.18***
PPE/A	L–H	606	–0.07	2.68	–7.99	15.12	0.60	5.51	0.24***	–0.09**
RD/A	H–L	516	0.25	2.28	–6.93	18.04	1.16	10.77	0.16*	–0.09**
BE/ME	H–L	606	0.54	2.62	–11.13	12.89	0.20	5.86	0.12**	0.03
EF/A	H–L	606	–0.48	1.44	–5.52	6.33	–0.01	4.50	0.11**	–0.08**
GS	H–L	606	–0.29	1.63	–6.01	4.78	–0.19	3.76	0.14***	0.00
BE/ME	L–M	606	–0.23	2.08	–10.53	14.03	0.33	10.05	0.09	–0.11***
EF/A	H–M	606	–0.29	1.66	–6.36	10.43	0.67	7.21	0.14**	–0.18***
GS	H–M	606	–0.23	1.81	–6.87	10.52	0.36	6.25	0.15**	–0.16***
BE/ME	H–M	606	0.33	1.52	–5.26	9.78	0.97	7.57	0.18***	–0.09**
EF/A	L–M	606	0.18	1.20	–2.98	6.06	0.90	5.45	0.13***	–0.15***
GS	L–M	606	0.06	1.68	–5.79	9.89	1.21	7.88	0.21***	–0.17***
<i>Panel B Sentiment proxies based on Baker–Wurgler sentiment</i>										
$\rho_{LR,t}$		604	–0.03	0.91	–2.08	2.09	–0.17	–0.19	0.99***	0.72***
$\eta_t - \eta_{t-1}$		601	0.00	0.02	–0.09	0.07	–0.24	1.35	0.15**	–0.26***
$(\rho_t - \rho_{t-1})^\perp$		601	0.00	0.02	–0.07	0.08	0.11	1.41	0.08	0.05
BN_LR		602	0.00	1.06	–2.32	3.34	0.28	3.67	0.99***	0.98***
BN_SR		602	0.00	0.22	–0.85	0.77	–0.41	4.64	0.91***	–0.26***

This table reports the descriptive statistics of dependent variables and the decomposed Baker–Wurgler sentiment measures. The first column is the variable used to sort portfolio. H is a portfolio of top three deciles; M is the middle four deciles; L contains the bottom three deciles. AR(1) represents the first-order autocorrelation. Corr represents the correlation between the summarised variable and one-term lagged Baker–Wurgler sentiment index. Panel A contains summary statistics for all the return premiums of the sixteen long-short strategies. In Panel A, the first column is the characteristic used and the second column is the portfolio construction. Panel B contains the statistics of long-run sentiment measure  $\rho_{LR,t}$  and two proxies for short-run sentiment.  $\eta_t - \eta_{t-1}$  is the changes of the sentiment deviation from long-run sentiment, and  $(\rho_t - \rho_{t-1})^\perp$  is the sentiment increment orthogonalized to the long-run sentiment over previous periods. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . The sample period is from July 1965 to September 2015

portfolio returns of ME(L–H) and BE/ME (H–L) have little autocorrelation, the returns of all other long-short portfolios are significantly autocorrelated. Furthermore, except for BE/ME(L–H), GS(H–L) and BE/ME(H–M), Baker–Wurgler sentiment negatively predicts future returns with strong statistical significance. The negative correlation coefficients between the long-short portfolio returns and 1-month lagged investor sentiment accord with previous studies showing that investor sentiment is a good contrarian predictor of future cross-sectional returns.

### 3.2 Decomposition of investor sentiment

To test our theoretical predictions on the cross-sectional effect of the long- and short-run sentiment, we empirically decompose the overall market sentiment, as measured by the original monthly sentiment index of Baker and Wurgler (2006), into long- and short-run components. We choose the Baker and Wurgler index as our baseline sentiment measure to make it easier to compare our results with theirs. Baker and Wurgler (2006) use the principal component analysis method to extract the common component of five sentiment proxies, including closed-end fund discount (CEFD), the number and the first-day returns of IPOs (NIPO, RIPO), the equity share in total new issues (S), and the dividend premium (P).<sup>15</sup> The Baker–Wurgler index, *Sent\_BW*, is orthogonalized to macroeconomic variables, including the growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment and the NBER dummy variable for recessions. The sample period is from July 1965 to September 2015. We also use other survey-based sentiment measures such as Consumer Confidence index to obtain the long- and short-run components and find similar results.

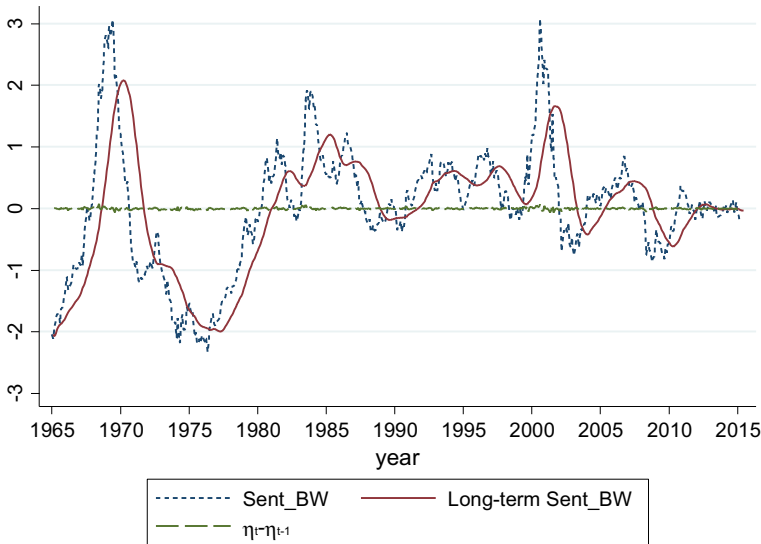
We implement two approaches to decompose the original investor sentiment proxy. The first one uses a moving average of the original sentiment proxy as a crude yet intuitive measure for the long-run sentiment component. More specifically, at each time  $t$ , the long-run sentiment component  $\rho_{LR,t}$  is the moving average of the original sentiment index over a 2-year period between  $[t - 25, t - 2]$ . While the choice of a 24-month window is admittedly somewhat arbitrary, it is partially motivated by the observation that periods of high/low sentiment often persist for around 2 years.<sup>16</sup>

When the long-run sentiment component is measured crudely by smoothing average, there are two ways to construct the corresponding short-run sentiment component. One measure of the short-run component  $(\rho_t - \rho_{t-1})^\perp$  is the change in the current sentiment from its previous level, which is also orthogonalized to the long-run sentiment component.  $\rho_t - \rho_{t-1}$  is orthogonalized from the long-run sentiment component to obtain a measure of the short-run sentiment fluctuation that is uncorrelated with the long run sentiment.

<sup>15</sup> We thank Jeffery Wurgler for providing these data. The Baker and Wurgler sentiment index used to be a common component of six proxies among which TURNOVER is included. Baker and Wurgler index no longer use TURNOVER due to the concern that TURNOVER does not represent sentiment well when high frequency trading is prevailing in the market.

<sup>16</sup> For example, the US stock market experienced a “new-issue mania” between 1961 and 1962, high investor sentiment for firms with strong growth potential between 1967 and 1968, and a bubble in gambling issues in 1977 and 1978. Concerning the bubble bursting, it also usually takes around 2 years for stock price to come back to earth in the anecdotal history. For instance, following the high-tech bubble in early 1980s, investors’ demand shifted to dividend paying stocks between 1987 and 1988. For robustness purposes, we also consider alternative windows of the moving average for long-run sentiment, including 12-month, 36-month and 48-month, and our conclusion remains unchanged.





**Fig. 1** Moving average based decomposition of BW Sentiment Index, 1965:07–2015:09. The dashed blue line depicts the original Baker–Wurgler sentiment index during July 1965 and September 2015. The solid red line depicts the long-run component of Baker–Wurgler sentiment index, which is measured by the moving average of previous 24-month sentiment. The green line depicts the short-run component  $\eta_t - \eta_{t-1}$ , which is the change in the deviation of current sentiment from its correspondent long-run sentiment (Color figure online)

Another measure for short-run sentiment  $\eta_t - \eta_{t-1}$  is the change in the deviation of current sentiment from its correspondent long-run sentiment  $(\rho_t - \rho_{LR,t}) - (\rho_{t-1} - \rho_{LR,t-1})$ .<sup>17</sup>

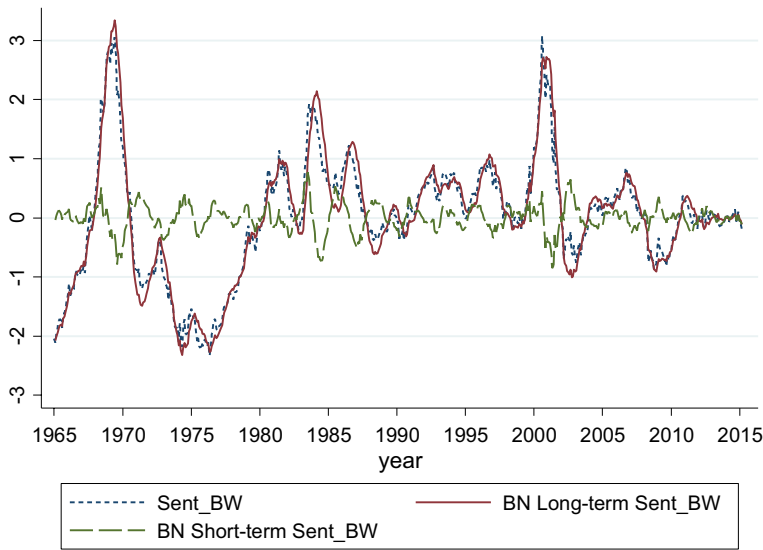
Our second approach to decompose sentiment is from Beveridge and Nelson (1981).<sup>18</sup> The Beveridge–Nelson decomposition is an approach to decompose the ARIMA(p, 1, q) process into two components: a permanent component that is a random walk with drift and a transitory component that is a stationary process with a mean of zero. We consider the permanent component of the decomposed sentiment index as the long-run sentiment (BN\_LR), and the transitory component of decomposed sentiment index as the short-run sentiment (BN\_SR).

Figure 1 depicts the time series of decomposed long- and short-run sentiment and the original Baker–Wurgler index when using a moving average to obtain long-run sentiment. The long-run sentiment is  $\rho_{LR}$  and the short-run sentiment is  $\eta_t - \eta_{t-1}$ . The graph shows that the long-run sentiment is strongly correlated with the original Baker–Wurgler sentiment index, albeit with some lags. The long-run sentiment is smoother than the original Baker–Wurgler index, while the short-run sentiment is relatively small and fluctuates around zero. The short-run sentiment component is generally smaller in magnitudes than the long-run sentiment component.

Figure 2 plots Beveridge–Nelson decomposed sentiment and the original Baker–Wurgler index. It shows that BN\_LR is highly correlated with the original Baker–Wurgler

<sup>17</sup> Based on our model, the short-run sentiment should be  $(\rho_t - (1 + r)\rho_{t-1})^\perp$ . We nevertheless follow the previous literature and ignore the effect of risky-free rate to obtain a short-run sentiment proxy,  $(\rho_t - \rho_{t-1})^\perp$ . We also use  $(\rho_t - (1 + r)\rho_{t-1})^\perp$  to run the tests and the regression results are strongly consistent with the results of using  $(\rho_t - \rho_{t-1})^\perp$ . The monthly risky-free rate is small and does not affect our main results.

<sup>18</sup> We thank Dominique Ladiray for providing the algorithm codes.



**Fig. 2** Beveridge–Nelson Decomposition of BW Sentiment Index, 1965:07–2015:09. The short-dashed blue line depicts the original Baker–Wurgler sentiment index. The red line depicts the long-run component of Baker–Wurgler sentiment index decomposed by Beveridge and Nelson (1981) method. The long-dashed green line depicts the short-run component of Baker–Wurgler sentiment index decomposed by Beveridge and Nelson (1981) method (Color figure online)

sentiment. Comparing Fig. 2 with Fig. 1, the long-run sentiment is no longer a lagged version of original sentiment. The correlation coefficient between the long-run sentiment and the original sentiment is higher when we use BN\_LR as the long-run sentiment indicator. Figure 2 also shows that BN\_SR has a broader range than other short-run sentiment measures, such as  $\eta_t - \eta_{t-1}$  and  $(\rho_t - \rho_{t-1})^\perp$ .

Panel B of Table 1 presents the descriptive statistics of the decomposed investor sentiment during the sample period from July 1965 to September 2015. In terms of magnitudes, the long-run sentiment is much larger than the short-run sentiment. The standard deviations of the long-run sentiment  $\rho_{LR}$  and BN\_LR are 0.91 and 1.06, respectively. The standard deviations of the two short-run sentiment components,  $\eta_t - \eta_{t-1}$  and  $(\rho_t - \rho_{t-1})^\perp$ , are both 0.02. The Beveridge–Nelson decomposition generates a short-run sentiment with relatively larger magnitude than  $\eta_t - \eta_{t-1}$  and  $(\rho_t - \rho_{t-1})^\perp$ . The short-run sentiment component, BN\_SR, has a standard deviation of 0.22.

Panel B also shows that the long-run sentiment measures, namely  $\rho_{LR}$  and BN\_LR, have significant first-order autocorrelation coefficient with a value of 0.99. Short-run sentiment measure  $(\rho_t - \rho_{t-1})^\perp$  does not have a significant correlation with its own lagged term, as it has been orthogonalized to the strongly persistent long-run sentiment component. The short-run sentiment BN\_SR is still significantly auto-correlated, with a first-order autocorrelation coefficient of 0.91. The last column of Panel B presents the correlation between each decomposed sentiment and the one-term lagged Baker–Wurgler sentiment. Apart from  $(\rho_t - \rho_{t-1})^\perp$ , the long- and short-run sentiment measures are significantly associated with the original sentiment, although the correlation coefficients for the short-run sentiment are relatively small in terms of magnitude. With the exception of  $(\rho_t - \rho_{t-1})^\perp$ , the short-term sentiment measures are negatively associated with the one-period lagged original Baker–Wurgler sentiment.

## 4 Empirical results

### 4.1 Cross-sectional returns on decomposed sentiment

Our theoretical model predicts that both the long- and short-run sentiment components affect cross-sectional stock returns. To test this prediction, we run the following regression:

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{LR,t} + \beta_2 \Delta \rho_{s,t} + \gamma X + u_t, \quad (27)$$

where  $R_{t,1} - R_{t,2}$  represents the relative returns of a sentiment-prone portfolio over a sentiment-immune portfolio,  $\rho_{LR,t}$  refers to the long-run sentiment component at time  $t$ ,  $\Delta \rho_{s,t}$  represents the short-run sentiment increments, and  $X$  is a vector of control variables. The control variables include Fama–French (2015) five factors (RMRF, SMB, HML, RMW, CMA) and the Carhart (1997) momentum factor (UMD).<sup>19</sup> Specifically, RMRF is the market return premium over risk-free rate; SMB is the average return on the three small portfolios minus the average return on the three big portfolios; HML is the average return on the two value portfolios minus the average return on the two growth portfolios; RMW is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios; CMA is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios; and UMD is the average return of high prior return portfolio over low prior return portfolio. The control variable SMB (HML) is excluded when the long-short portfolio is constructed with ME (BE/ME). The control variable RMW is excluded when the long-short portfolio is constructed with E/BE or D/BE.

The strong persistence of the long-run sentiment measure may raise the concern of spurious regressions. Stambaugh (1999) points out that the coefficient estimates of predictive regression with a small sample can be biased and distort the  $t$ -statistics when the predictor is highly persistent. Stambaugh bias exists if the autoregressive disturbance of a lagged stochastic regressor correlates with the regression error term. Under this circumstance, OLS regression results will lead to an erroneous conclusion that the lagged regressors have predictive power while in fact they do not. Thus, it is paramount that we account for Stambaugh bias in our predictive regressions. To this end, we adopt the multi-predictor augmented regression method of Amihud et al. (2009) to adjust for the Stambaugh bias in the estimated coefficients and report the  $t$ -statistics of coefficients estimated from a wild bootstrap procedure. The detailed methodology of this wild bootstrap procedure is in the “Appendix”. We also calculate Newey–West standard errors (Newey and West 1986) that are robust to heteroscedasticity and serial correlation, and we choose a maximum lag of 12 throughout the regressions.

Table 2 reports the estimation coefficients on the standardized long- and short-run sentiment components of the Baker–Wurgler sentiment indicator. All independent variables (including the long- and short-run sentiment measures) are standardised before running the regression to facilitate comparisons of coefficients. The short-run sentiment in Panel A ( $\eta_t - \eta_{t-1}$ ) is measured by changes in the sentiment deviation from the long-run sentiment. The short-run sentiment in Panel B,  $(\rho_t - \rho_{t-1})^\perp$ , is measured by the sentiment increment orthogonalized to the long-run sentiment component. The long-run sentiment component ( $\rho_{LR,t}$ ) in both Panels A and B is the moving average of prior  $[-25, -2]$  monthly investor sentiment. The long- and short-run sentiment components in Panel C are Beveridge–Nelson decomposed long- and short-run sentiment, BN\_LR

<sup>19</sup> The data are available on [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We thank Kenneth R. French for providing the data.

**Table 2** Regressions of monthly cross-sectional returns on decomposed sentiment

		Panel A		Panel B		Panel C	
		$\rho_{LR,t}$	$\eta_t - \eta_{t-1}$	$\rho_{LR,t}$	$(\rho_t - \rho_{t-1})^\perp$	BN_LR	BN_SR
ME	L-H	-0.262*** (0.000)	0.332*** (0.000)	-0.323*** (0.000)	0.312*** (0.000)	-0.236*** (0.000)	0.135*** (0.000)
Age	L-H	-0.017 (0.126)	0.187*** (0.000)	-0.051*** (0.000)	0.175*** (0.000)	-0.091*** (0.000)	0.109*** (0.000)
Sigma	H-L	-0.195*** (0.000)	0.147*** (0.000)	-0.222*** (0.000)	0.138*** (0.000)	-0.242*** (0.000)	-0.001 (0.273)
E/BE	<0->0	-0.335*** (0.000)	0.180*** (0.000)	-0.368*** (0.000)	0.170*** (0.000)	-0.453*** (0.000)	0.039*** (0.000)
D/BE	=0->0	-0.305*** (0.000)	0.059*** (0.000)	-0.316*** (0.000)	0.056*** (0.000)	-0.359*** (0.000)	-0.101*** (0.000)
PPE/A	L-H	0.060*** (0.000)	0.014 (0.377)	0.058*** (0.000)	0.013 (0.310)	0.062*** (0.000)	0.164*** (0.000)
RD/A	H-L	-0.032*** (0.002)	0.010** (0.038)	-0.033*** (0.000)	0.009** (0.040)	-0.007*** (0.000)	-0.119*** (0.000)
BE/ME	H-L	-0.084*** (0.000)	0.116*** (0.000)	-0.105*** (0.000)	0.109*** (0.000)	-0.104*** (0.000)	-0.004 (0.188)
EF/A	H-L	0.002*** (0.002)	-0.003 (0.136)	0.003** (0.011)	-0.003** (0.035)	0.007*** (0.002)	-0.031*** (0.000)
GS	H-L	-0.052*** (0.000)	-0.127*** (0.000)	-0.029*** (0.000)	-0.119*** (0.000)	0.020*** (0.000)	-0.092*** (0.000)
BE/ME	L-M	-0.004*** (0.001)	-0.060*** (0.000)	0.007*** (0.000)	-0.057*** (0.000)	-0.016*** (0.005)	-0.002 (0.114)
EF/A	H-M	-0.103*** (0.000)	0.008 (0.335)	-0.105*** (0.000)	0.008 (0.195)	-0.102*** (0.000)	-0.024*** (0.000)
GS	H-M	-0.114*** (0.000)	0.020*** (0.000)	-0.118*** (0.000)	0.019*** (0.000)	-0.093*** (0.000)	-0.029*** (0.000)
BE/ME	H-M	-0.088*** (0.000)	0.056*** (0.000)	-0.098*** (0.000)	0.052*** (0.000)	-0.120*** (0.000)	-0.006 (0.419)
EF/A	L-M	-0.105*** (0.000)	0.010*** (0.000)	-0.107*** (0.000)	0.010*** (0.000)	-0.109*** (0.000)	0.008*** (0.000)
GS	L-M	-0.062*** (0.000)	0.147*** (0.000)	-0.089*** (0.000)	0.138*** (0.000)	-0.112*** (0.000)	0.063*** (0.000)

This table reports the regressions of long-short portfolio returns on both the long-run and short-run sentiment

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{LR,t} + \beta_2 \Delta \rho_{s,t} + \gamma X + u_t$$

$R_{t,1} - R_{t,2}$  represents the return disparity of more sentiment-prone portfolio over the less sentiment-prone portfolio. The control variables (X) include the Fama–French Five factors (RMRF, HML, SMB, RMW, CMA), and the momentum factor (UMD). SMB (HML) will not be included in regression when return premium is constructed by ME (BE/ME). The first two columns show how the portfolio is constructed. H, M, L represents the top three, middle four and bottom three decile portfolios respectively. The long-run sentiment component  $\rho_{LR,t}$  in Panel A and Panel B is the standardized smoothing average of prior  $[-25, -2]$  monthly investor sentiment. Short term component in Panel A and Panel B are respectively the standardized incremental change of sentiment deviation from long-run sentiment average  $\eta_t - \eta_{t-1}$  and the standardized incremental sentiment orthogonalized to long-run sentiment  $(\rho_t - \rho_{t-1})^\perp$ . Panel C presents the effect of BW sentiment decomposed with Beveridge and Nelson (1981) method; the long- and short-run sentiment are noted as BN\_LR and BN\_SR respectively. All coefficients are adjusted for Stambaugh-bias. The  $p$  values reported in parentheses from wild bootstrap procedures in which all stimulation uses Newey West robust t-statistics. See “Appendix” for details of bootstrap simulation. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

and BN\_SR, respectively. We report the Stambaugh bias-adjusted coefficients with bootstrapped  $p$  values in parentheses.

Panel A in Table 2 shows that the long-run sentiment component  $\rho_{LR,t}$  is a significant contrarian predictor for 13 out of 16 long-short portfolio returns, consistent with our theory. Furthermore, the short-run sentiment component  $\eta_t - \eta_{t-1}$  is significantly and positively associated with 11 out of 16 long-short portfolio returns, which is consistent with existing evidence that contemporaneous excess returns are positively related to shifts in sentiment (Lee et al. 2002; Ben-Rephael et al. 2012).

In general, the empirical results in Panel A provide strong support for our two main hypotheses. The signs of coefficients on both the long- and short-run sentiment components in 11 out of 16 regressions are consistent with the predictions of our theory. However, the coefficients on the long-run sentiment are significantly positive for the regressions involving PPE/A(L–H) and EF/A(H–L), while the coefficients on the short-run sentiment are significantly negative in the case of GS(H–L) and BE/ME(L–H). These portfolios with the unexpected signs of the regression coefficients are portfolios in the “Tangibility” and in “Growth Opportunity and Distress” group. Baker and Wurgler (2006) also find that investor sentiment is not a good predictor of the future returns of the portfolios in the “Tangibility” group and argue that the multidimensional nature of BE/ME, EF/A, and GS makes the results unclear in the “Growth Opportunity and Distress” group.

Panel B reports the estimation results of the long-run sentiment  $\rho_{LR,t}$  and the short-run sentiment  $(\rho_t - \rho_{t-1})^\perp$ . The coefficients on the long-run sentiment component are significantly negative in 13 of the 16 regressions and the coefficients on the short-run sentiment component are significantly positive in 11 out of 16 regressions, consistent with the results in Panel A. The coefficients on the short-run sentiment are significantly positive in 5 out of the 16 regressions. The dependent variables in those regressions with unexpected coefficients on the long run and short sentiment belong to the “Tangibility” and the “Growth Opportunity and Distress” group. The magnitude of the coefficients on the long- and the short-run sentiment measures in Panel B is in most cases comparable to their counterparts in Panel A.

Panel C reports the estimation results of the Beveridge–Nelson decomposed long- and short-run sentiment. The coefficients on the long-run sentiment BN\_LR are almost all significantly negative, while those on the short-run sentiment BN\_SR are significantly positive in six out of the 16 regressions. Under the Beveridge–Nelson decomposition, the long-run sentiment no longer exhibits a pattern lagged to original sentiment (recall that in Fig. 2), and yet is still negatively associated with the subsequent long-short portfolio returns. Despite some differences in the magnitude, the coefficients on the long- and short-run sentiment components in Panel C have the same signs as their counterparts in Panel A.

Regressions with the long-short portfolio returns as the dependent variables may obscure the effects of the two sentiment components on individual decile portfolios. To address this issue, we run regressions of decile portfolio returns on both the long- and short-run sentiment variables and control variables:

$$R_{t,i,j} = \alpha + \beta_{1,i,j}\rho_{LR,t} + \beta_{2,i,j}(\eta_t - \eta_{t-1}) + \gamma X + u_t. \quad (28)$$

$R_{t,i,j}$  represents the return of the  $i$ th decile portfolio sorted by variable  $j$ , where  $i$  represents the decile portfolio rank and takes values from 1 to 10 and  $j$  is one of the ten firm characteristic variables used to construct the decile portfolios. The control variables ( $X$ ) include the Fama–French five factors (RMRF, HML, SMB, RMW, CMA), and the momentum factor (UMD).

Table 3 reports the coefficients on the long- and short-run sentiment components. Panel A documents a large variation in the coefficients on the long-run sentiment component

across the decile portfolios. The observed patterns are consistent with predictions of our theoretical model. Specifically, we find decile portfolios that are more prone to market-wide sentiment are usually more affected by the long run sentiment. More specifically, for Size and Age sorted portfolios, the coefficients of the long-run sentiment increase from Decile 1 to Decile 10 most of the time, indicating that the reversal effect of long-run sentiment effect on decile returns is larger on small and young stocks than large and old stocks. The coefficients on the long-run sentiment for Sigma sorted decile portfolios decrease almost monotonically with the decile rank, implying that high long-run sentiment leads to lower returns on more volatile decile portfolios than less volatile portfolios. For the two variables in the “Tangibility” group, PPE/A and RD/A, there is no clear pattern in the long-run sentiment coefficients across different deciles. This finding is consistent with Baker and Wurgler (2006) and is in line with the results in Table 2.

For firm characteristics with the multidimensional nature, namely BE/ME, EF/A, and GS, we find that the coefficients on the long-run sentiment across deciles have an inverse U-shape. This indicates that the middle deciles are less prone to overall market, while distressed stocks and stocks with strong growth potential are more prone to overall market. When sorted by ME, Sigma, and D/BE, the less sentiment-prone deciles and the more sentiment-prone deciles show different exposure to the long-run sentiment. For example, for deciles sorted on Sigma, the coefficients of the bottom two deciles are significantly positive, indicating that the “bond-like” stocks with low return volatility have negative exposure to the long-run overall market sentiment. One plausible explanation proposed by Baker and Wurgler (2007) is “flights to quality”. When overall market is pessimistic, bond-like stocks are more appealing to not only sophisticated investors but also noise traders, leading to rising prices of those stocks during low market sentiment periods. The reverse pattern we found in “bond-like” stocks helps explain the weak relationship between investor sentiment and aggregate market return documented in the literature. Indeed, in an untabulated regression, we test the effects of decomposed sentiment on aggregate market returns and find the coefficients of both the long- and short-run sentiment components are insignificant.

Panel B of Table 3 reports the coefficients of the short-run sentiment component. The coefficients on the short-term sentiment are positive and significant in almost all more sentiment-prone decile portfolios. There is a decreasing (increasing) pattern in the coefficients of the short-run sentiment across the deciles sorted by ME, Age, E/BE and D/BE (Sigma). However, the coefficients on the short-run sentiment follow a U-shaped pattern across the deciles sorted by BE/ME, EF/A, and GS and exhibit no clear pattern in the deciles sorted by PPE/A and RD/A. We also find significantly negative coefficients on the short-run sentiment in most of the less sentiment-prone deciles. In general, the results in Panel B confirm the conclusions drawn from Panel A that the effect of short-run sentiment on returns varies across deciles and the bond-like stocks have negative exposure to overall market sentiment.

## 4.2 Robustness checks

This paper presents a behavioural explanation for the variations in the cross-sectional stock returns. However, it is possible that variation in investor sentiment reflects changes in systematic risk and our results may not be entirely consistent with the behavioural story. For example, changes in our decomposed sentiment measures may coincide with time variation in the market beta. If that is the case, the cross-sectional patterns conditional on the ten characteristics would represent a rational compensation for systematic risk. We test this possibility with a time-varying CAPM beta model.



**Table 3** Decile portfolio returns and decomposed sentiment

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
<i>Panel A Coefficients of long-run sentiment component <math>\beta_{1,i,j}</math></i>										
ME	-0.115*** (0.000)	-0.105*** (0.000)	-0.027*** (0.000)	-0.046*** (0.000)	-0.088*** (0.000)	-0.095*** (0.000)	0.040*** <sub>a</sub> (0.000)	-0.068*** (0.000)	-0.026*** (0.000)	-0.067*** (0.000)
Age	-0.116*** (0.000)	-0.033*** (0.000)	0.027*** <sub>a</sub> (0.000)	0.092*** <sub>a</sub> (0.000)	0.119*** <sub>a</sub> (0.000)	-0.054*** (0.000)	-0.013*** (0.000)	0.025*** <sub>a</sub> (0.000)	-0.100*** (0.000)	-0.174*** (0.000)
Sigma	0.138*** <sub>a</sub> (0.000)	0.058*** <sub>a</sub> (0.000)	-0.009*** (0.000)	-0.024*** (0.000)	0.018*** <sub>a</sub> (0.000)	-0.071*** (0.000)	-0.023*** (0.000)	-0.097*** (0.000)	-0.139*** (0.000)	-0.158*** (0.000)
E/BE	-0.126*** (0.000)	-0.124*** (0.000)	-0.011*** (0.000)	-0.007*** (0.000)	-0.020*** (0.000)	-0.031*** (0.000)	0.029*** <sub>a</sub> (0.000)	-0.068*** (0.000)	-0.113*** (0.000)	-0.140*** (0.000)
D/BE	-0.212*** (0.000)	-0.183*** (0.000)	-0.131*** (0.000)	-0.070*** (0.000)	-0.118*** (0.000)	-0.051*** (0.000)	0.042*** (0.000)	-0.295*** (0.000)	-0.244*** (0.000)	0.061*** <sub>a</sub> (0.000)
PPE/A	-0.100*** (0.000)	-0.029*** (0.000)	-0.011* (0.070)	-0.117*** (0.000)	-0.055*** (0.000)	-0.096*** (0.000)	-0.161*** (0.000)	-0.084*** (0.000)	-0.084*** (0.000)	-0.165*** (0.000)
RD/A	-0.121*** (0.000)	-0.326*** (0.000)	-0.333*** (0.000)	-0.207*** (0.000)	-0.252*** (0.000)	-0.200*** (0.000)	-0.042*** (0.000)	-0.122*** (0.000)	-0.085*** (0.000)	-0.024*** (0.000)
BE/ME	-0.117*** (0.000)	-0.092*** (0.000)	0.043*** <sub>a</sub> (0.000)	0.000*** <sub>a</sub> (0.032)	-0.001*** (0.000)	0.070*** <sub>a</sub> (0.000)	0.007 (0.376)	-0.006*** (0.006)	-0.059*** (0.000)	-0.005 (0.243)
EF/A	-0.170*** (0.000)	-0.083*** (0.000)	0.032*** <sub>a</sub> (0.000)	0.016*** <sub>a</sub> (0.000)	0.044*** <sub>a</sub> (0.000)	0.014*** <sub>a</sub> (0.000)	0.043*** <sub>a</sub> (0.000)	-0.013*** (0.000)	-0.039*** (0.000)	-0.169*** (0.000)
GS	-0.171*** (0.000)	0.038*** <sub>a</sub> (0.000)	-0.008*** (0.000)	0.062*** <sub>a</sub> (0.000)	0.015*** <sub>a</sub> (0.000)	-0.042*** (0.000)	0.001*** <sub>a</sub> (0.000)	-0.090*** (0.000)	-0.068*** (0.000)	-0.153*** (0.000)
<i>Panel B Coefficients of short-run sentiment component <math>\beta_{2,i,j}</math></i>										
ME	0.272*** (0.000)	-0.054*** <sub>a</sub> (0.000)	-0.012 (0.415)	-0.119*** <sub>a</sub> (0.000)	-0.081*** <sub>a</sub> (0.000)	-0.152*** <sub>a</sub> (0.000)	0.004* (0.099)	-0.028*** <sub>a</sub> (0.000)	0.023*** (0.000)	-0.064*** <sub>a</sub> (0.000)
Age	0.165*** (0.000)	0.162*** (0.000)	0.057*** (0.000)	0.028*** (0.000)	0.104*** (0.000)	0.058*** (0.000)	0.028*** (0.000)	-0.036*** <sub>a</sub> (0.000)	-0.027*** <sub>a</sub> (0.000)	-0.016*** <sub>a</sub> (0.000)
Sigma	0.024*** (0.000)	0.009*** (0.000)	-0.013*** <sub>a</sub> (0.000)	0.024*** (0.000)	0.022*** (0.000)	-0.009*** <sub>a</sub> (0.000)	0.069*** (0.000)	0.039*** (0.000)	0.113*** (0.000)	0.304*** (0.000)

**Table 3** (continued)

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
E/BE	0.250*** (0.000)	0.017*** (0.000)	0.027*** (0.000)	−0.040*** <sub>a</sub> (0.000)	−0.003 (0.119)	−0.048*** <sub>a</sub> (0.000)	−0.031*** <sub>a</sub> (0.000)	0.004** (0.003)	−0.216*** <sub>a</sub> (0.000)	−0.019*** <sub>a</sub> (0.000)
D/BE	0.124*** (0.000)	−0.017*** <sub>a</sub> (0.001)	−0.054*** <sub>a</sub> (0.000)	−0.012*** <sub>a</sub> (0.000)	−0.088*** <sub>a</sub> (0.000)	−0.144*** <sub>a</sub> (0.000)	0.025*** (0.000)	−0.060*** (0.000)	0.011*** (0.000)	0.067*** (0.000)
PPE/A	0.094*** (0.000)	0.073*** (0.000)	0.106*** (0.000)	0.128*** (0.000)	0.093*** (0.000)	0.043*** (0.000)	0.073*** (0.000)	0.062*** (0.000)	0.080*** (0.000)	0.082*** (0.000)
RD/A	0.085*** (0.000)	−0.027*** <sub>a</sub> (0.000)	0.059*** (0.000)	0.096*** (0.000)	−0.078*** <sub>a</sub> (0.000)	0.006** (0.050)	0.070*** (0.000)	0.136*** (0.000)	0.277*** (0.000)	0.268*** (0.000)
BE/ME	0.136*** (0.000)	0.131*** (0.000)	0.071*** (0.000)	0.063*** (0.000)	0.173*** (0.000)	0.101*** (0.000)	0.058*** (0.000)	0.114*** (0.000)	0.109*** (0.000)	0.126*** (0.000)
EF/A	0.129*** (0.000)	0.050*** (0.000)	0.102*** (0.000)	0.104*** (0.000)	0.051*** (0.000)	0.082*** (0.000)	0.084*** (0.000)	0.014*** (0.000)	0.077*** (0.000)	0.183*** (0.000)
GS	0.236*** (0.000)	0.211*** (0.000)	0.103*** (0.000)	0.042*** (0.000)	0.083*** (0.000)	−0.024*** <sub>a</sub> (0.000)	0.043*** (0.000)	0.064*** (0.000)	−0.030*** <sub>a</sub> (0.000)	0.117*** (0.000)

This table reports the regressions of long-short portfolio returns on both the long-run and short-run sentiment

$$R_{t,i,j} = \alpha + \beta_{1,i,j}\rho_{LR,t} + \beta_{2,i,j}(\eta_t - \eta_{t-1}) + \gamma X + u_t$$

$R_{t,i,j}$  represents the return of the  $i$ th decile portfolio sorted by variable  $j$ , where  $i$  is from 1 to 10 and  $j$  is one of the ten sentiment-prone characteristic variables. The control variables ( $X$ ) include the Fama–French Five factors (RMRF, HML, SMB, RMW, CMA), and the momentum factor (UMD). SMB (HML) is not included in regression when the long-short portfolio is constructed by ME (BE/ME). Long-run sentiment component  $\rho_{t-1}^*$  is the standardized smoothing average of prior  $[-25, -2]$  monthly investor sentiment, and short term component is the standardized incremental change of sentiment deviation from long-run sentiment average  $\eta_t - \eta_{t-1}$ . Panel A and Panel B respectively reports the coefficients long- and short-run sentiment. The label “a” indicates that the sign of the coefficient is contrary to the hypothesis. The coefficients are adjusted for Stambaugh-bias. The  $p$  values are obtained from wild bootstrap procedures in which all stimulation uses Newey West robust  $t$ -statistics. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{LR,t} + \beta_2 \Delta \rho_{s,t} + (b + \gamma_1 \rho_{LR,t} + \gamma_2 \Delta \rho_{s,t}) RMRF_t + u_t, \quad (29)$$

where  $R_{t,1} - R_{t,2}$  represents the portfolio returns that long the more sentiment-prone portfolios and short the less sentiment-prone portfolios,  $\rho_{LR,t}$  refers to the long-run sentiment component at time  $t$ ,  $\Delta \rho_s$  represents the short-run sentiment increments, and  $RMRF_t$  is the market return premium. If the negative (positive) effect of long-run (short-run) sentiment on the cross-sectional return is driven by its negative (positive) effect on the beta loading of market return premium, the coefficients for the interaction terms will be significantly different from zero, and the sign of  $\gamma_i$  will be the same as the sign of  $\beta_i$  in Table 2; otherwise, the behavioural story holds. That is to say,  $\gamma_1$  should be significantly negative and  $\gamma_2$  should be significantly positive if the rational explanation holds.

Table 4 shows the sign and magnitude of the coefficients on both the long- and short-run sentiment components remain consistent with their counterparts in Table 2 even after including the interaction terms in the regressions. This evidence suggests that the long- and short-run sentiment components do indeed affect the cross-sectional stock returns. We also consider another potential systematic risk explanation, which posits that even when the market beta is constant, the decomposed investor sentiment may reflect the variations in the market return premium. If this story holds, the decomposed investor sentiment should perform well in predicting the market return premium. However, in an unreported regression of market return premium on decomposed investor sentiment, we find little evidence that decomposed investor sentiment components affect the aggregate market returns.

One may also argue that the effect of investor sentiment may come from the cross-sectional variation in investor attention. As several studies show that rising investor attention leads to predictable returns both in time-series and in the cross-section (e.g., Barber and Odean 2008; Da et al. 2015; Peng and Xiong 2006; Li and Yu 2012), it remains unclear whether stocks that appear to be most affected by investor sentiment are actually those that attract most attention from investors. To shed some light on this issue, we investigate the correlation between investor sentiment and the difference between attention of more sentiment-prone stocks and less sentiment-prone stocks. We use two investor attention measures proposed by Barber and Odean (2008), i.e., the abnormal trading volume index and the abnormal return index. We first calculate the monthly abnormal trading volume and abnormal return indexes for each firm and then calculate the average investor attention for each portfolio. The abnormal trading volume (return) index is calculated as the ratio of the stock's trading volume (return) in a given month to its average trading volume (return) over the prior 1-year. We construct the cross-sectional attention disparity in the same way as we calculate the return disparity of the sentiment-prone portfolio over the sentiment-immune portfolio. Take ME sorted long-short portfolio as an example, the attention disparity is the difference in the average attention between the bottom three deciles and the top three deciles of size sorted portfolios. In an unreported test, we find that the correlation coefficients between attention disparity and the change in market sentiment are usually relatively small and mostly insignificant.

We also examine whether the long- and short-run sentiment components lose their explanatory and predictive power after adding the contemporaneous cross-sectional investor attention disparity measure as an additional control variable in our regressions. The regression results are shown in Table 5. Panel A of Table 5 presents the regression results when investor attention  $A_{t,1}$  is measured by abnormal trading volume and Panel B shows the results when attention disparity is measured by abnormal return. The results suggest that the investor attention disparity is significantly related to cross-sectional stock returns. However, the effects of the decomposed sentiment components on returns remain strong

**Table 4** Time-varying market betas

		Panel A				Panel B			
		$\rho_{LR,t}$	$\eta_t - \eta_{t-1}$	$\rho_{LR,t}RMRF_t$	$(\eta_t - \eta_{t-1})RMRF_t$	$\rho_{LR,t}$	$(\rho_t - \rho_{t-1})^\perp$	$\rho_{LR,t}RMRF_t$	$(\rho_t - \rho_{t-1})^\perp RMRF_t$
ME	L-H	-0.218*** (0.000)	0.301** (0.046)	-3.122*** <sub>a</sub> (0.008)	0.005 (0.367)	-0.340*** (0.000)	0.233*** (0.000)	-3.110*** <sub>a</sub> (0.004)	0.007 (1.000)
Age	L-H	-0.634*** (0.000)	0.232*** (0.004)	0.961* (0.081)	0.187 (0.162)	-0.732*** (0.000)	0.143*** (0.003)	0.946 (0.187)	0.182*** <sub>a</sub> (0.000)
Sigma	H-L	-1.507*** (0.000)	0.206 (0.462)	0.133*** (0.000)	0.293*** <sub>a</sub> (0.001)	-1.587*** (0.000)	0.107*** (0.000)	0.090 (0.316)	0.285 (1.000)
E/BE	<0->0	-1.108*** (0.000)	0.367 (0.385)	2.419*** (0.000)	0.453*** <sub>a</sub> (0.000)	-1.255*** (0.000)	0.242*** (0.000)	2.394 (0.153)	0.447 (1.000)
D/BE	=0->0	-1.094*** (0.000)	0.116 (0.316)	2.062** (0.034)	0.242*** <sub>a</sub> (0.050)	-1.143*** (0.000)	0.015*** (0.000)	2.004 (0.123)	0.231*** <sub>a</sub> (0.000)
PPE/A	L-H	-0.285*** (0.000)	-0.031* (0.096)	0.271*** (0.000)	0.187*** <sub>a</sub> (0.008)	-0.273*** (0.000)	-0.030*** (0.002)	0.263*** (0.000)	0.185*** <sub>a</sub> (0.000)
RD/A	H-L	-0.401*** (0.000)	-0.165*** (0.000)	2.419*** (0.000)	0.047 (0.234)	-0.341*** (0.000)	-0.219*** (0.000)	2.341*** (0.002)	0.032*** <sub>a</sub> (0.000)
BE/ME	H-L	0.176*** (0.000)	0.275* (0.084)	-3.766* (0.078)	-0.060 (0.212)	0.061*** (0.000)	0.286*** (0.000)	-3.696*** <sub>a</sub> (0.045)	-0.048*** (0.000)
EF/A	H-L	-0.145*** (0.000)	-0.078*** (0.000)	1.827** (0.010)	-0.005 (0.172)	-0.114*** (0.000)	-0.104 (0.472)	1.790*** (0.000)	-0.012*** (0.000)
GS	H-L	-0.125*** (0.000)	-0.247** (0.019)	0.400 (0.131)	-0.099* (0.074)	-0.023*** (0.000)	-0.239*** (0.000)	0.357*** (0.000)	-0.107*** (0.000)
BE/ME	L-M	-0.294*** (0.000)	-0.123*** (0.000)	1.506*** (0.010)	0.096 (0.255)	-0.243*** (0.000)	-0.167** (0.023)	1.446 (0.293)	0.086*** <sub>a</sub> (0.000)
EF/A	H-M	-0.420*** (0.000)	0.009*** (0.000)	1.598* (0.094)	0.088*** <sub>a</sub> (0.000)	-0.426*** (0.000)	-0.039 (0.376)	1.559 (0.420)	0.081*** <sub>a</sub> (0.000)

**Table 4** (continued)

		Panel A				Panel B			
		$\rho_{LR,t}$	$\eta_t - \eta_{t-1}$	$\rho_{LR,t}RMRF_t$	$(\eta_t - \eta_{t-1})RMRF_t$	$\rho_{LR,t}$	$(\rho_t - \rho_{t-1})^\perp$	$\rho_{LR,t}RMRF_t$	$(\rho_t - \rho_{t-1})^\perp RMRF_t$
GS	H–M	−0.455*** (0.000)	−0.027*** (0.007)	2.256*** (0.000)	0.026 (0.122)	−0.444*** (0.000)	−0.082*** (0.000)	2.207*** (0.000)	0.016*** <sub>a</sub> (0.000)
BE/ME	H–M	−0.253*** (0.000)	0.138 (0.228)	−2.236 (0.375)	0.047 (0.143)	−0.310*** (0.000)	0.109 (0.200)	−2.225*** <sub>a</sub> (0.000)	0.048*** <sub>a</sub> (0.000)
EF/A	L–M	−0.227*** (0.000)	0.090*** (0.000)	−0.348*** <sub>a</sub> (0.001)	0.082*** <sub>a</sub> (0.000)	−0.262*** (0.000)	0.063 (0.243)	−0.349 (0.261)	0.081*** <sub>a</sub> (0.004)
GS	L–M	−0.521*** (0.000)	0.214 (0.241)	1.972** (0.029)	0.134*** <sub>a</sub> (0.000)	−0.611*** (0.000)	0.157** (0.017)	1.973*** (0.006)	0.133*** <sub>a</sub> (0.000)

This table reports regressions of long-short portfolio returns on the market risk premium and its interactions with both long-run sentiment and short-run sentiment

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{LR,t} + \beta_2 \Delta \rho_{s,t} + (b + \gamma_1 \rho_{LR,t} + \gamma_2 \Delta \rho_{s,t}) RMRF_t + u_t$$

$R_{t,1} - R_{t,2}$  the return disparity of more sentiment-prone portfolio over the less sentiment-prone portfolio. The long-short portfolios are constructed based on the firm characteristics following Baker and Wurgler (2006) approach. The Baker–Wurgler sentiment index is decomposed. The long-run sentiment component  $\rho_{LR,t}$  is the standardized smoothing average of prior  $[-25, -2]$  monthly investor sentiment. Short-run sentiment component  $\Delta \rho_{s,t}$  is measured by  $\eta_t - \eta_{t-1}$  and  $\rho_t - \rho_{t-1}$  in Panel A and Panel B respectively. All regressors are standardized. The coefficients are adjusted for Stambaugh-bias. The label “a” marks the coefficients of interaction terms significantly supports the rational story. The  $p$  values reported in parentheses are obtained from wild bootstrap procedures in which all stimulation uses Newey West robust t-statistics. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

even after controlling for investor attention. The results in Panels A and B are almost the identical in terms of the sign and significance of the coefficients for the long- and the short-run sentiment. As predicted, we show that the long-run sentiment negatively predicts future cross-sectional returns and the short-run sentiment is positively correlated with contemporaneous cross-sectional return premium. Although unreported, including lagged investor attention disparity measures does not change our results either. Therefore we conclude that cross-sectional investor attention is unlikely to drive our results.

It is commonly known that measures of investor sentiment are noisy (Baker and Wurgler 2006). While the Baker–Wurgler sentiment index has become a workhorse sentiment indicator, it relies on variables such as returns, volatility, and trading volume, which are themselves equilibrium outcomes. Another way to measure sentiment is to use survey data. Following prior literature (e.g., Lemmon and Portniaguina 2006; Ho and Hung 2009), we also use two survey-based sentiment indicators, the Consumer Confidence Index by Michigan (ICS) and the Conference Board Consumer Confidence Index (CCI) from Bloomberg, as alternative sentiment measures. We report the results in Table 6. Panel A shows that the long-run sentiment exhibit a significantly negative effect on future return in 9 out of 16 the cases, with 8 of these 9 cases having a significantly positive coefficients on the short-run sentiment. Most of the cases that do not fit our expectation are related to BE/ME-, EF/A-, and GS-based portfolios. For these portfolios, the results are also mixed in Baker and Wurgler's (2006) study. The results in Panel B are largely consistent with their counterparts in Panel A, especially when looking into the coefficients of long-run sentiment. Thus, overall, the results of the survey-based sentiment indicators are generally consistent with those from decomposed Baker–Wurgler Sentiment.

We also conduct some other robustness checks (the results are not reported for the sake of brevity). First, we re-run our regressions with decomposed sentiment indicators from other widely accepted investor sentiment proxies, such as the closed-end fund discount (CEFD) proposed by Lee et al. (1991) and the aligned sentiment indicator (Sent\_PLS) from Huang et al. (2015). We find similar results when using decomposed CEFD and Sent\_PLS in the regression, especially for the coefficients of the long-run sentiment. Second, we construct different measures of sentiment components by taking the moving average of the Baker–Wurgler sentiment index over different horizons as the measure of the long-run sentiment. Specifically, our conclusions remain unchanged when the 12-month and 36-month horizon are used to calculate long-run sentiment. Third, we divide the samples into high and low sentiment periods, where a low (high) sentiment period is defined as the period when the current sentiment is lower (higher) than the previous 2-year smoothing average sentiment. We find that the long- and short-run sentiment components perform better at explaining the cross-sectional returns during periods of high sentiment. This evidence may be explained by more binding short-selling constraints during high sentiment periods (e.g., Nagel 2005; Stambaugh et al. 2012; Yu and Yuan 2011). Fourth, because of the close association of investor sentiment premium and macroeconomics announcement (Du and Hu 2018), we include a set of additional macroeconomic control variables (dividend-price ratio, price-earnings ratio, term spread, and default spread) in our regressions. Furthermore, we include the liquidity factor and bid-ask spreads to account for the effect of liquidity. The results are consistent with the baseline analysis. Fifth, it might be argued that the financial and firm characteristics used to construct our long-short portfolios are noisy proxies of the degree of exposure to market-wide sentiment. To address this concern, we use principal component analysis (PCA) to build an index of exposure to investor sentiment based on these accounting and financial variables, and test whether the decomposed sentiment can explain or predict the index. Using PCA helps remove the idiosyncratic noise



**Table 5** Effects of decomposed investor sentiment after controlling for investor attention

		Panel A Investor attention measured by abnormal trading volume			Panel B Investor attention measured by abnormal return		
		$\rho_{t-1}^*$	$\eta_t - \eta_{t-1}$	$A_{t-1,1} - A_{t-1,2}$	$\rho_{t-1}^*$	$\eta_t - \eta_{t-1}$	$A_{t-1,1} - A_{t-1,2}$
ME	L-H	-0.125*** (0.000)	0.273*** (0.000)	6.407*** (0.000)	-0.246*** (0.000)	0.331*** (0.000)	0.002*** (0.000)
Age	L-H	0.049*** (0.000)	0.156*** (0.000)	2.658*** (0.000)	-0.012** (0.045)	0.191*** (0.000)	0.003*** (0.000)
Sigma	H-L	-0.088*** (0.000)	0.124*** (0.000)	5.456*** (0.000)	-0.174*** (0.000)	0.153*** (0.000)	-0.003*** (0.000)
E/BE	<0->0	-0.167*** (0.000)	0.162*** (0.000)	-0.720*** (0.000)	-0.317*** (0.000)	0.176*** (0.000)	-0.001*** (0.000)
D/BE	=0->0	-0.157*** (0.000)	0.055*** (0.000)	4.653*** (0.000)	-0.297*** (0.000)	0.060*** (0.000)	-0.001*** (0.000)
PPE/A	L-H	0.090*** (0.000)	0.004 (0.250)	7.708*** (0.000)	0.091*** (0.000)	0.014*** (0.006)	-0.001*** (0.000)
RD/A	H-L	0.024*** (0.000)	-0.019*** (0.000)	5.011*** (0.000)	-0.033*** (0.000)	0.013*** (0.000)	0.000 (0.135)
BE/ME	H-L	0.004 (0.254)	0.078*** (0.000)	3.965*** (0.000)	-0.070*** (0.000)	0.115*** (0.000)	-0.000*** (0.000)
EF/A	H-L	-0.023*** (0.000)	-0.012*** (0.000)	5.893*** (0.000)	-0.000** (0.030)	-0.002 (0.115)	-0.001*** (0.000)
GS	H-L	-0.082*** (0.000)	-0.129*** (0.000)	2.821*** (0.000)	-0.053*** (0.000)	-0.131*** (0.000)	-0.002*** (0.000)
BE/ME	L-M	0.016*** (0.000)	-0.060*** (0.000)	4.838*** (0.000)	-0.003* (0.072)	-0.060*** (0.000)	-0.002*** (0.000)
EF/A	H-M	-0.098*** (0.000)	0.005** (0.047)	6.207*** (0.000)	-0.106*** (0.000)	0.006*** (0.002)	-0.001*** (0.000)
GS	H-M	-0.095*** (0.000)	-0.003*** (0.000)	6.174*** (0.000)	-0.115*** (0.000)	0.020*** (0.000)	-0.002*** (0.000)
BE/ME	H-M	0.020*** (0.000)	0.016*** (0.000)	4.268*** (0.000)	-0.070*** (0.000)	0.055*** (0.000)	0.001*** (0.000)
EF/A	L-M	-0.057*** (0.000)	0.023*** (0.000)	4.706*** (0.000)	-0.103*** (0.000)	0.009*** (0.000)	-0.000 (0.102)
GS	L-M	-0.008*** (0.009)	0.128*** (0.000)	2.345*** (0.000)	-0.061*** (0.000)	0.152*** (0.000)	-0.004*** (0.000)

This table reports the regressions of long-short portfolio returns on both the long-run and short-run sentiment

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{t-1}^* + \beta_2 (\eta_t - \eta_{t-1}) + \beta_3 (A_{t,1} - A_{t,2}) + \gamma X + u_t$$

$R_{t,1} - R_{t,2}$  ( $A_{t,1} - A_{t,2}$ ) represents the return (investor attention) disparity of more sentiment-prone portfolio over the less sentiment-prone portfolio. The control variables (X) include the Fama-French Five factors (RMRF, HML, SMB, RMW, CMA), and the momentum factor (UMD). The first two columns show how the portfolio is constructed. H, M, L represents the top three, middle four and bottom three decile portfolios respectively. The long-run sentiment component  $\rho_{t-1}^*$  is the standardized smoothing average of prior  $[-25, -2]$  monthly investor sentiment.  $\eta_t - \eta_{t-1}$  is the standardized incremental change of sentiment deviation from long-run sentiment average. Panel A (B) reports the results when  $A_t$  is measured by abnormal trading volume (abnormal return). The  $p$  values reported in parentheses are obtained from wild bootstrap procedures in which all stimulation uses Newey West robust t-statistics. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

**Table 6** Regressions of monthly cross-sectional returns on survey-based sentiment indicators

		Panel A ICS		Panel B CCI	
		$\rho_{t-1}^*$	$\eta_t - \eta_{t-1}$	$\rho_{t-1}^*$	$\eta_t - \eta_{t-1}$
ME	L-H	-0.17*** (0.000)	0.75 (0.188)	-0.16*** (0.000)	0.49 (0.154)
Age	L-H	0.01*** (0.000)	0.31* (0.056)	0.07*** (0.000)	0.00** (0.013)
Sigma	H-L	-0.40*** (0.000)	0.18*** (0.000)	-0.04*** (0.000)	-0.05*** (0.000)
E/BE	<0->0	-0.12*** (0.000)	0.32 (0.254)	-0.58 (0.309)	-0.06** (0.030)
D/BE	=0->0	-0.29*** (0.000)	0.30** (0.046)	-0.29*** (0.000)	-0.07*** (0.000)
PPE/A	L-H	-0.03*** (0.000)	0.17*** (0.000)	-0.02** (0.016)	-0.04*** (0.000)
RD/A	H-L	0.22*** (0.000)	0.00*** (0.000)	0.22*** (0.000)	-0.06*** (0.000)
BE/ME	H-L	-0.01*** (0.000)	0.15*** (0.000)	-0.11*** (0.000)	0.24*** (0.001)
EF/A	H-L	0.14*** (0.000)	0.00 (0.137)	0.16*** (0.000)	-0.13 (0.126)
GS	H-L	0.10*** (0.000)	-0.08 (0.295)	0.22*** (0.000)	-0.22*** (0.001)
BE/ME	L-M	0.04*** (0.000)	0.03** (0.034)	0.02*** (0.000)	-0.08 (0.184)
EF/A	H-M	-0.01*** (0.000)	0.10** (0.011)	0.02*** (0.000)	-0.05*** (0.000)
GS	H-M	0.09*** (0.000)	0.02* (0.079)	0.18*** (0.000)	-0.07*** (0.000)
BE/ME	H-M	0.03** (0.014)	0.17*** (0.001)	-0.11*** (0.000)	0.15 (0.216)
EF/A	L-M	-0.13*** (0.000)	0.10 (0.241)	-0.18*** (0.000)	0.08*** (0.004)
GS	L-M	-0.04*** (0.000)	0.10* (0.057)	-0.07*** (0.000)	0.15*** (0.006)

This table reports the regressions of long-short portfolio returns on both the long-run and short-run sentiment

$$R_{t,1} - R_{t,2} = \alpha + \beta_1 \rho_{t-1}^* + \beta_2 (\eta_t - \eta_{t-1}) + \gamma X + u_t$$

$R_{t,1} - R_{t,2}$  represents the return disparity of more sentiment-prone portfolio over the less sentiment-prone portfolio. The control variables (X) include the Fama-French Five factors (RMRF, HML, SMB, RMW, CMA), and the momentum factor (UMD). The first two columns show how the portfolio is constructed. H, M, L represents the top three, middle four and bottom three decile portfolios respectively. The long-run sentiment component  $\rho_{t-1}^*$  is the standardized smoothing average of prior  $[-25, -2]$  monthly investor sentiment.  $\eta_t - \eta_{t-1}$  is the standardized incremental change of sentiment deviation from long-run sentiment average. Panel A reports the effect of decomposed Consumer Confidence Index by Michigan (ICS), and Panel B presents the effect of decomposed Conference Board Consumer Confidence Index

**Table 6** (continued)

(CCI). The  $p$  values reported in parentheses are obtained from wild bootstrap procedures in which all stimulation uses Newey West robust t-statistics. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

embedded in the individual accounting and financial variables. We find our results continue to hold. Finally, our conclusions remain unchanged when we re-run our tests using value-weighted portfolio returns to isolate the size effect on the portfolio returns.

## 5 Conclusion

This study uses a simple and straightforward model to show that investor sentiment affects returns of different assets disproportionately. We extend Delong et al. (1990) model to a noise trader risk model with multiple risky assets. In our model, we allow the risky assets to have different exposure to overall market investor sentiment and provide theoretical predictions that are consistent with the empirical evidence of the effect of investor sentiment on the cross-sectional stock returns.

Motivated by the model, we also decompose investor sentiment into the long- and short-run components. Consistent with the theory, we find that the long-run sentiment component is a contrarian predictor of future long-short portfolio returns and the short-run sentiment is positively correlated with contemporaneous long-short portfolio returns, where the long-short portfolios long sentiment-prone stocks and short sentiment-immune stocks. Furthermore, we test whether the effect of the sentiment components can be attributed to the time-varying beta loading of the market premium (or other risk factors). We show that the effect of sentiment components on the cross-sectional return is not related to systematic risk. Accordingly, the behavioural story holds. Further analysis suggests that our results are robust to alternative sentiment measures, different sample periods, additional control variables, and the use of value-weighted returns.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix 1: Detailed description of variables representing sentiment-prone level

Table 7 gives a detailed description of the variables needed to construct the portfolios.

**Table 7** Definitions of characteristic variables for long-short strategies

Var	Name	Description	Calculation
ME	Market equity	Price times shares outstanding in June before year t. If there is more than one permanent code for a company, then sum up all the ME for the same company	$\text{abs}(\text{prc}) * \text{shrout}$
Age	Firm age	The number of months between the firm's first appearance on CRSP and the year t. Measure Age to the nearest month. If the stock is not delisted, we calculate period length between current year t and beginning date, or else the age is ending date minus beginning date	$\text{min}(\text{date}, \text{enddat}) - \text{begdat}$
$\sigma$	Total risk	Annual standard deviation in monthly returns from CRSP for the 12 months ending in the June before year t, and there should be no less than 9 monthly returns available to estimate it	$\text{std}(\text{retadj})$
E/BE	Earnings-book ratio for profitable firms	Earnings are income before extraordinary items (Item 18) plus income statement deferred taxes (Item 50) minus preferred dividends (Item 19). If earnings are positive; book equity (BE) is shareholders' equity (Item 60) plus balance sheet deferred taxes (Item 35). The profitability dummy E > 0	$\text{BE} = \text{CEQ} + \text{TXDITC}$ ; $\text{E} = \text{IB} + \text{TXDI} - \text{DVP}$ ; $\text{E/BE} = \text{E/BE}$ if $\text{E} > 0$ ; $\text{E/BE} = 0$ if $\text{E} < 0$
D/BE	Dividend-book ratio for dividend payers	Dividend is the fiscal year-end dividends per share at the ex-date (Item 26) times Compustat shares outstanding (Item 25) divided by book equity	$\text{D/BE} = (\text{DVPSX\_F} * \text{CSHO}) / \text{BE}$ if $\text{D} > 0$ ; otherwise $\text{D/BE} = 0$
PPE/A	Fixed assets ratio	Plant, Property, and Equipment (Item 7) is scaled by total gross assets (Item 6). The data are widely available after 1971. We do not replace missing value with zero	$\text{PPE/A} = \text{PPEGT/AT}$
RD/A	Research and development ratio	Research and Development (Item 46) is also scaled by total gross assets (Item 6). The data are widely available after 1971	$\text{RD/A} = \text{XRD/AT}$
BE/ME	Book-to-market ratio	This is the log of the ratio of book equity to market equity. We match fiscal year ending calendar year t – 1 ME with June t BE	$\log(1 + \text{BE/DEC\_ME})$

**Table 7** (continued)

Var	Name	Description	Calculation
EF/A	External finance over assets	External finance (EF) is equal to the change in assets (Item 6) less the change in retained earnings (Item 36). When the change in retained earnings is not available, we use net income (Item 172) less common dividends (Item 21) instead	$EF1 = \text{dif}(\text{RE}); EF2 = \text{dif}(\text{NI-DVC}); EF/A = (\text{dif}(\text{AT}) - \text{coalesce}(EF1, EF2, 0)) / \text{AT}$
GS	Sales growth	Sales growth is the percentage change in net sales (Item 12). We first calculate the original sales growth ratio and then use its position in the ten-decile to note GS. GS has a range from [1, 10]	$GS = \text{dif}(\text{SALE}) / \text{lag}(\text{SALE})$

## Appendix 2: Details of the wild bootstrap procedures to obtain $P$ values in regression tables

The null of this wild bootstrap  $p$  values is that the independent variable has no predictability. Take the regression in Table 2 as an example. The regression function is

$$R_t = \alpha + \beta_1 \rho_{LR,t} + \beta_2 \Delta \rho_{s,t} + \gamma X + u_t,$$

where  $R_t$  is the relative returns of more sentiment-prone stocks over less sentiment-prone stocks ( $R_t = R_{t,1} - R_{t,2}$ ).

To obtain the simulated data of dependent variable, we first run an OLS regression of the original regression function to get the fitted residuals  $\hat{\epsilon}_t$ .

$$\hat{\epsilon}_t = R_t - (\hat{\alpha} + \hat{\beta}_1 \rho_{LR,t} + \hat{\beta}_2 \Delta \rho_{s,t} + \hat{\gamma} X)$$

Assuming the predictors follow an AR(1) process, to obtain the simulated sample of independent variables, we run first-order autoregression and get the fitted residuals of the AR(1) regression for each predictor.

$$\hat{\phi}_{\rho_{LR,t}} = \rho_{LR,t} - (\hat{\delta}_{t-1,0} + \hat{\delta}_{t-1,1} \rho_{LR,t-1})$$

Then we generate a set of random number,  $\omega_t$ , from the standard normal distribution. We build up a pseudo sample of observations for relative returns and the variables that have no return predictability under the null hypothesis.

$$\tilde{R}_t = \bar{R}_t + \hat{\epsilon}_t \omega_t,$$

$$\widetilde{\rho_{LR,t}} = (\hat{\delta}_{t-1,0} + \hat{\delta}_{t-1,1} \rho_{LR,t-1}) + \hat{\phi}_{\rho_{LR,t}} \omega_t,$$

where  $\bar{R}_t$  is the sample mean of  $R_t$ , and  $\omega_t$  is a drawn from a standard normal distribution.

With the pseudo sample, we estimate the coefficients and the corresponding Newey–West t-statistics proposed by Newey and West (1986) for each regressor. We then repeat this process for 500 times and store all the Newey–West robust t-statistics for each regressor. We get a distribution of the bootstrapped t-statistics for each regressor.

Because our model suggests a negative sign of  $\beta_1$  and a positive sign of  $\beta_2$ , we test null hypotheses  $H_0: \beta_1 = 0$  and  $\beta_2 = 0$  against alternative hypotheses  $H_A: \beta_1 < 0$  and  $\beta_2 > 0$ . The one-sided tests are more in line with our theory than the two-sided tests are. Our results remain significant when we use two-sided tests instead. For a given regressor, the empirical  $p$  value is the proportion of the bootstrapped t-statistics larger (smaller) than the t-statistics when using the original sample.

## References

- Aboody D, Even-Tov O, Lehavy R, Trueman B (2018) Overnight returns and firm-specific investor sentiment. *J Financ Quant Anal* 53:485–505
- Amihud Y, Hurvich CM, Wang Y (2009) Multiple-predictor regressions: hypothesis testing. *Rev Financ Stud* 22:413–434
- Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. *J Finance* 61:1645–1680
- Baker M, Wurgler J (2007) Investor sentiment in the stock market. *J Econ Perspect* 21:129–151



- Baker M, Wurgler J, Yuan Y (2012) Global, local, and contagious investor sentiment. *J Financ Econ* 104:272–287
- Barber BM, Odean T (2008) All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Rev Financ Stud* 21:785–818
- Barberis N, Shleifer A, Vishny R (1998) A model of investor sentiment. *J Financ Econ* 49:307–343
- Ben-Rephael A, Kandel S, Wohl A (2012) Measuring investor sentiment with mutual fund flows. *J Financ Econ* 104:363–382
- Berger D, Turtle H (2012) Cross-sectional performance and investor sentiment in a multiple risk factor model. *J Bank Finance* 36:1107–1121
- Beveridge S, Nelson CR (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. *J Monetary Econ* 7:151–174
- Bhattacharya U, Galpin N, Ray R, Yu X (2009) The role of the media in the internet IPO bubble. *J Financ Quant Anal* 44:657–682
- Brown GW, Cliff MT (2004) Investor sentiment and the near-term stock market. *J Empir Finance* 11:1–27
- Brown GW, Cliff MT (2005) Investor sentiment and asset valuation. *J Bus* 78:405–440
- Campbell JY, Kyle AS (1993) Smart Money, Noise trading and stock price behaviour. *Rev Econ Stud* 60:1–34
- Carhart MM (1997) On persistence in mutual fund performance. *J Finance* 52:57–82
- Chen C, Kuo I (2014) Investor sentiment and interest rate volatility smile: evidence from Eurodollar options markets. *Rev Quant Financ Acc* 43:367–391
- Chen N-F, Kan R, Miller MH (1993) Are the discounts on closed-end funds a sentiment index? *J Finance* 48:795–800
- Cochrane JH, Longstaff FA, Santa-Clara P (2008) Two trees. *Rev Financ Stud* 21:347–385
- Daniel K, Hirshleifer D, Subrahmanyam A (1998) Investor psychology and security market under- and over-reactions. *J Finance* 53:1839–1885
- Da Z, Engelberg J, Gao P (2015) The sum of all FEARS investor sentiment and asset prices. *Rev Financ Stud* 28:1–32
- Delong JB, Shleifer A, Summers LH, Waldmann RJ (1990) Noise trader risk in financial markets. *J Polit Econ* 98:703–738
- Du D, Hu O (2018) The sentiment premium and macroeconomic announcements. *Rev Quant Financ Acc* 50:207–237
- Elton EJ, Gruber MJ, Busse JA (1998) Do investors care about sentiment? *J Bus* 71:477–500
- Fama EF, French KR (1992) The cross-section of expected stock returns. *J Finance* 47:427–465
- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116:1–22
- Frazzini A, Lamont OA (2008) Dumb money: mutual fund flows and the cross-section of stock returns. *J Financ Econ* 88:299–322
- Hirshleifer D (2001) Investor psychology and asset pricing. *J Finance* 56:1533–1597
- Ho C, Hung CH (2009) Investor sentiment as conditioning information in asset pricing. *J Bank Finance* 33:892–903
- Huang D, Jiang F, Tu J, Zhou G (2015) Investor sentiment aligned: a powerful predictor of stock returns. *Rev Financ Stud* 28:791–837
- Hutton P, Lee F, Shu Z (2012) Do managers always know better? The relative accuracy of management and analyst forecasts. *J Acc Res* 50:1217–1244
- Jackson AB, Rountree BR, Sivaramakrishnan K (2017) Earnings co-movements and earnings manipulation. *Rev Acc Stud* 22:1340–1365
- Jackson AB, Plumlee MA, Rountree BR (2018) Decomposing the market, industry, and firm components of profitability: implications for forecasts of profitability. *Rev Acc Stud* 23:1071–1095
- Kim SH, Kim D (2014) Investor sentiment from internet message postings and the predictability of stock returns. *J Econ Behav Organ* 107:708–729
- Kumar A, Lee CMC (2006) Retail investor sentiment and return comovements. *J Finance* 61:2451–2486
- Lee CMC, Shleifer A, Thaler RH (1991) Investor sentiment and the closed-end fund puzzle. *J Finance* 46:75–109
- Lee WY, Jiang CX, Indro DC (2002) Stock market volatility, excess returns, and the role of investor sentiment. *J Bank Finance* 26:2277–2299
- Lemmon M, Portniaguina E (2006) Consumer confidence and asset prices: some empirical evidence. *Rev Financ Stud* 19:1499–1529
- Li C-A, Yeh C-C (2011) Investor psychological and behavioral bias: do high sentiment and momentum exist in the china stock market? *Rev Pac Basin Financ Mark Polic* 14:429–448

- Li J, Yu J (2012) Investor attention, psychological anchors, and stock return predictability. *J Financ Econ* 104:401–419
- Lucas RE (1978) Asset prices in an exchange economy. *Econometrica* 46:1429
- Nagel S (2005) Short sales, institutional investors and the cross-section of stock returns. *J Financ Econ* 78:277–309
- Newey WK, West KD (1986) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. NBER, Massachusetts, USA
- Peng L, Xiong W (2006) Investor attention, overconfidence and category learning. *J Financ Econ* 80:563–602
- Qiu L, Welch I (2004) Investor sentiment measures. NBER, Working Paper
- Schmeling M (2009) Investor sentiment and stock returns: some international evidence. *J Empir Finance* 16:394–408
- Sharma SS, Narayan P, Thuraisamy K (2015) Time-varying herding behavior, global financial crisis, and the Chinese stock market. *Rev Pac Basin Financ Mark Polic*. <https://doi.org/10.1142/S0219091515500095>
- Shleifer A, Vishny RW (1997) The limits of arbitrage. *J Finance* 52:35–55
- Stambaugh RF (1999) Predictive regressions. *J Financ Econ* 54:375–421
- Stambaugh RF, Yu JF, Yuan Y (2012) The short of it: investor sentiment and anomalies. *J Financ Econ* 104:288–302
- Tetlock PC (2007) Giving content to investor sentiment: the role of media in the stock market. *J Finance* 62:1139–1168
- Verrecchia RE (2001) Essays on disclosure. *J Acc Econ* 32:97–180
- Yu J, Yuan Y (2011) Investor sentiment and the mean-variance relation. *J Financ Econ* 100:367–381